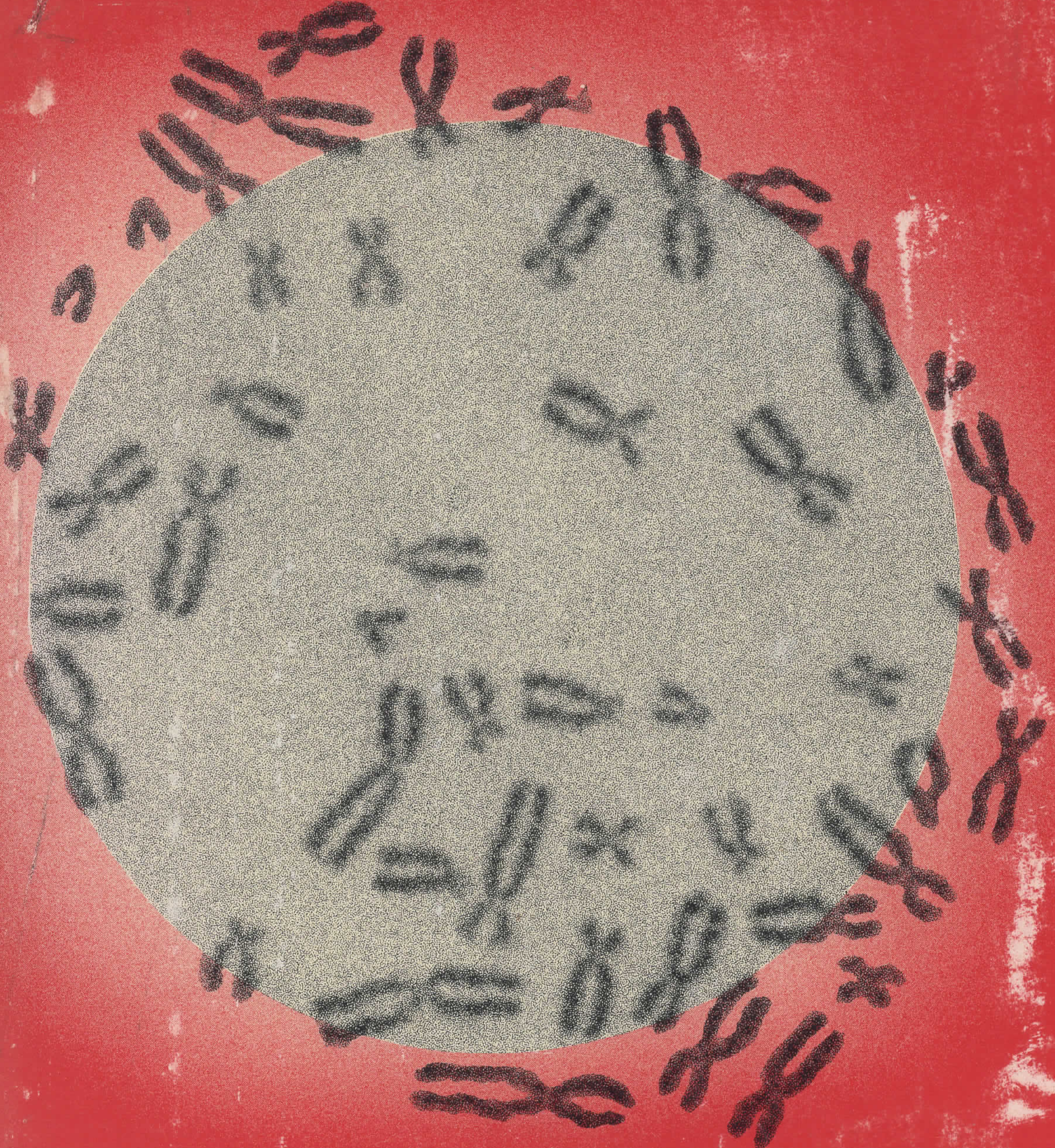


Genetics Notes

James F. Crow

Seventh Edition

Robert V. Miller

GENETICS NOTES

by

James F. Crow

University of Wisconsin
Madison, Wisconsin

Seventh Edition

LIBRARY OF ROBERT V. MILLER

Burgess Publishing Company
Minneapolis, Minnesota

CONTENTS

Foreword v

1 Mendelian Inheritance 1
2 The Chemical Nature of the Gene 11
3 Chromosome Behavior and the Chromosomal Basis of Heredity 19
4 Sex-Chromosomes and Sex-Linkage 27
5 Probability 39
6 Analysis of Human Pedigrees 45
7 Life Cycles 51
8 Linkage and Chromosome Mapping 57
9 Bacterial and Viral Genetics 69
10 Gene Interaction and the Effect of the Environment 77
11 Biochemical Genetics 89
12 DNA, RNA, and Protein 95
13 The Genetic Code 105
14 Multiple Alleles and Genetic Fine Structure 111
15 Chromosome Changes 121
16 Mutation 133
17 Cytoplasmic Heredity 141
18 Regulation of Gene Action 145
19 Immunogenetics 151
20 Population Genetics 161
21 Inbreeding 169
22 Selection 181
23 Quantitative Inheritance 193
24 Statistical Analysis of Quantitative Characters 203
25 Genetics and Evolution 213
26 The Origin of Life 225

FOREWORD

The first edition of these Notes was written exactly 25 years ago. It has been an incredible quarter century for genetics. At that time microbial genetics was in its infancy. The gene had not been proven to be DNA, although there was emerging evidence. In the first edition I alluded to the forthcoming chemical attack on the gene, but no one (at least, not I) dreamed that the attack would be so quickly successful.

The unfolding story of the chemistry of the gene was reflected in successive editions. Through the sixth edition, the structure of DNA and how this so elegantly explains the central mysteries of genetics was kept secret and not revealed until late in the book, after the classical groundwork had been laid. But it was increasingly apparent that the secret had been widely leaked, so I have now started the DNA story near the beginning.

This edition is somewhat longer than previous ones have been. Partly this reflects the increase of knowledge in molecular genetics. But I have also expanded the sections on population, quantitative, and evolutionary genetics, since these topics seem to me to be the least adequately dealt with in most genetics textbooks.

These Notes are intended to be self-contained, but they are a bit thin—skeleton thin in some parts. The pages have wide margins so you can write notes, clarifications, and additional material. I have given references, mostly to books that are more complete and detailed. These are assembled by chapter at the end of the book.

Most of the chapters are followed by problems, so you have a chance to test your understanding of the material and your ability to apply this knowledge. There is a glossary, an index, and a statistical appendix. Also at the end, there are answers to the problems. I mention this especially, since one unusually systematic student studied the book diligently, starting at the beginning and working all the problems, and didn't discover the answers until he had reached the end.

The chromosome photographs in Chapters 3, 4, and 15 were provided by Klaus and Eeva Patau. I thank many colleagues and other friends, too numerous to mention individually, for comments and suggestions. Robert DeMars and Millard Susman have been particularly helpful. My special thanks go to students who in various ways, direct and indirect, have revealed those parts which were vague, ambiguous, or simply wrong. Since so much of this edition has been rewritten, I again need your help.

J. F. C.

January, 1976

ts who in
rts which
ch of this

J. F. C.

1

MENDELIAN INHERITANCE

GENETICS BEFORE 1900

Eighteenth-century plant hybridizers, especially Kolreuter who published his results in the 1760's, were familiar with the pollination process, but there was considerable uncertainty as to whether only one pollen grain was sufficient for fertilization. But by the time of Mendel's work there was enough evidence that he developed his analysis in the correct belief that a single egg and pollen unite to initiate embryonic development.

Similarly, it was known that in animals a single egg and sperm unite. The early breeders were also aware that, despite the great disparity in size between the male and female germ cells, they usually contribute equally to the inheritance.

The cell theory was well established by Mendel's time, but the nuclear details of cell division were vague. The period between the publication of Mendel's paper in 1865 and its rediscovery in 1900 was one of great advance in the knowledge of chromosome behavior, so that when Mendel's principles became known it was immediately realized that the chromosomes provided the material basis for these principles. It was also during this period that Miescher discovered nucleic acid, although the identification of DNA as the chemical basis of heredity was not firmly established until the 1950's.

The prevailing view of heredity in the pre-Mendelian era was a **blending** theory. The hereditary material was thought of as being analogous to a fluid. Under this concept the progeny of a black and a white animal would be uniformly gray. The further progeny from crossing the hybrids among themselves would also be gray; for the black and white hereditary material, once blended, could never be separated again—just as a mixture of black and white paint cannot be separated.

Mendel's great contribution was to replace the blending theory with a **particulate** theory. Mendel first presented his rules of inheritance in 1865, but they remained unknown for many years. Their rediscovery in 1900 by deVries, Correns, and von Tschermak, independently in three different countries, marks the beginning of modern genetics.

MENDEL'S EXPERIMENTS

Mendel chose to work with garden peas because they were readily hybridized and the characters could be easily observed.

Seven character differences were selected for study: round vs. wrinkled peas, yellow vs. green peas, tall (6 feet) vs. short (1 foot) plants, and four others.

When he crossed a tall with a short variety the hybrids were all tall (regardless of which type was used as female parent and which the male, confirming observations of earlier workers). Then he allowed the hybrids from each cross to be self-pollinated and observed the subsequent generation. He found 787 tall and 277 short plants, roughly 3:1.

When the plants of this generation were self-fertilized as was done in the previous generation, Mendel found that all the short plants produced only short offspring. On the other hand, the tall plants were of two types. About 1/3 of them produced only tall offspring; the remaining 2/3 produced both tall and short plants in the proportions of roughly 3/4 and 1/4.

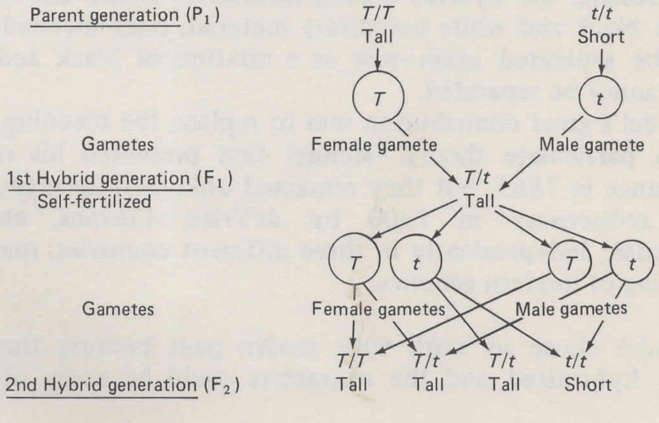
The same results occurred in crosses involving the other six characters.

MENDEL'S HYPOTHESIS AND HOW IT WAS TESTED (Mendel's First Law)

From the results just described, Mendel was able to formulate the remarkable hypothesis that has come to be known as **Mendel's First Law**, or the **Law of Segregation**. The principle is that heredity characters (such as tallness or shortness) are determined by particulate **factors** (now called genes); that these factors occur in pairs; and that in the formation of the gametes these factors are segregated so that only one of the pair is transmitted by a particular gamete. When the male and female gametes fuse to form the zygote, the double number is restored.

In addition to the principle of segregation, Mendel discovered **dominance**. He noted that the hybrid between the tall and short strain was tall, even though it must have contained both a tall and a short factor. Because the tall character seemed to dominate in the hybrid, Mendel designated the factor for tallness as **dominant** and the alternative **recessive**.

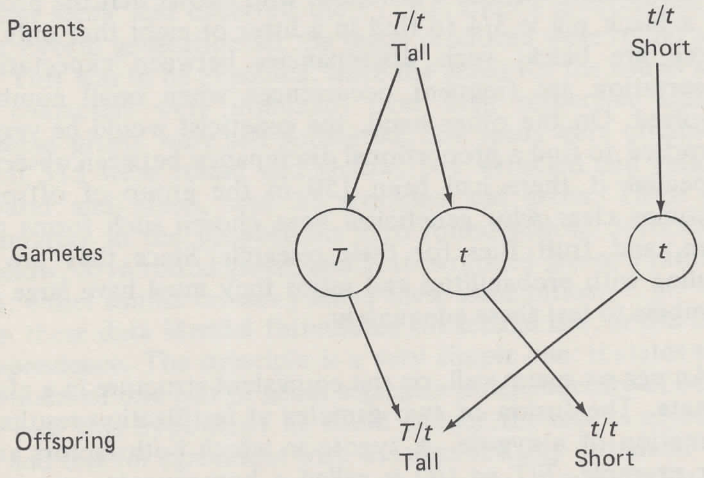
It is a convention, started by Mendel, to designate dominant factors with capital letters and recessives with corresponding lowercase letters. Using these symbols, the crosses between tall and short plants may be diagrammed as follows:



By combining the two kinds of male and female gametes in all possible ways, the relative proportions of offspring types are obtained. The hypothesis leads to an expectation of 3/4 tall and 1/4 short plants, a ratio of 3:1. Among the tall plants, 1/3 have two *T* factors, and hence if self-fertilized would produce only tall offspring; the other 2/3 have one *T* and one *t* factor and when self-fertilized would be expected to produce both tall and short progeny. So the theory accounts for all the results described in the previous section.

Mendel's law is not a specification of the physical nature of the character determiners but is essentially a statement of the rules they follow. The hypothesis can be tested by seeing if the rules are followed in further crosses. That is what Mendel did. He computed the proportion of tall and short plants to be expected from types of crosses other than those already done and compared the expectations with observed results from experimental tests.

The first test make was what is today called a **testcross**. Mendel crossed the F_1 hybrid (T/t) back to the short parent (t/t). This cross may be diagrammed as follows:



The expectation in this cross is ½ tall and ½ short. The actual results were 87 tall and 79 short. Here, as in all the other tests, the results were as close to a 1:1 ratio as would be expected according to the laws of chance.

PROBABILITY AND MENDELIAN EXPECTATIONS

According to the Mendelian hypothesis, the gametes are combined by the rules of chance with certain fixed probabilities. The situation is analogous to that of tossing coins or throwing dice. Common experience has demonstrated that the larger the number of trials, the more uniform are the results. If half a dozen coins are tossed, it is not an especially unlikely event to have them all come up heads; but if a large number were tossed, it is almost certain that both heads and tails would be found in approximately equal proportions.

The same thing is true of Mendelian class frequencies. When a ratio of 3/4 to 1/4 is expected in a group of half a dozen animals, almost anything might happen, and it does in practice. On the other hand, with large numbers of offspring, the numbers become more and more accurately predictable. One reason for Mendel's success was that he used large numbers of experimental plants and hence got easily recognizable ratios.

In many situations that one meets in genetics the probability of future events is predictable. From Mendel's hypothesis it can be stated that the probability is 3/4 that a given F_2 plant will be tall, or that the probability is 1/4 that it will be short. Likewise, in certain human families a geneticist can say that the probability is 1/2, or 1/4, that a child will be brown eyed, or be color-blind, or have some other trait whose manner of inheritance is known. Mathematical methods have been developed to determine the probability of a given discrepancy between theoretical and observed ratios. The practicing geneticist makes a great deal of use of such theory in testing hypotheses, and some procedures will be taken up later.

It does not surprise a geneticist who knows that the probability of a black pig is 3/4 to find in a litter of eight that only two or three are black; such discrepancies between expectation and observation are frequent occurrences when small numbers are involved. On the other hand, the geneticist would be very much surprised to find a proportional discrepancy between observed and expected if there had been 150 in the group of offspring. It becomes clear why geneticists have chosen such forms as peas, corn, and fruit flies for their research. Since they are always dealing with probabilities and ratios they must have large enough numbers to test these adequately.

SOME DEFINITIONS

An egg or sperm cell, or the equivalent structure in a plant, is a **gamete**. The fusion of two gametes at fertilization results in the formation of a **zygote**. A zygote in which both factors are alike (for example, *T*/*T* or *t*/*t*) is called a **homozygote**; one in which they are different is called a **heterozygote** (for example, *T*/*t*). The adjectives **homozygous** and **heterozygous** are used to describe these two conditions.

The hereditary factors are now called **genes**. The gene that shows its trait in the heterozygous condition is the dominant factor; that which is concealed is the recessive factor. Two such factors forming a contrasting pair are called **alleles**. The genetic makeup (*T*/*T*, *T*/*t*, or *t*/*t*) is the **genotype**. The character determined by the genetic makeup is the **phenotype**. For example, the two classes *T*/*T* and *T*/*t*, both being tall, have the same phenotype, but different genotypes.

ABSENCE OF DOMINANCE

Since the time of Mendel's experiments many examples have appeared where the heterozygote is intermediate between the two homozygotes, rather than like one or the other as would be expected if the character were completely dominant. A conspicuous example is the coat color in Shorthorn cattle. The genotypes and phenotypes are as follows:

Genotypes	Phenotype
R/R	Solid color (Red in Shorthorns)
R/r	Mottled (Roan)
r/r	White

The probability of the occurrence of various kinds of offspring of two roan parents, as determined by the usual method, would be: red, 1/4; white, 1/4; roan, 1/2. In a very large family these expected proportions would be approximately realized.

FURTHER EXPERIMENTS OF MENDEL
(Mendel's Second Law)

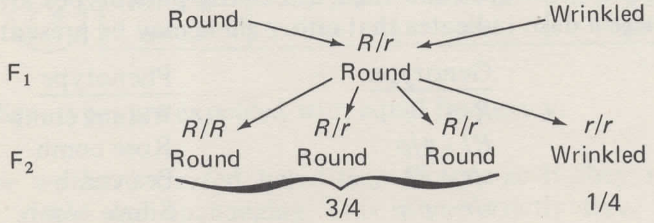
Mendel's first law, the law of segregation, applies to the behavior of a single pair of factors. The second law describes the simultaneous behavior of two or more pairs of genes.

Mendel made several hybridizations in which two or three pairs of factors were studied simultaneously. For example, he crossed plants producing round yellow seeds with plants producing wrinkled green seeds and observed the F_1, F_2, and F_3 generations. In the first hybrid generation all the seeds produced were round and yellow. This was to be expected, since the genes for the round and yellow characters were dominant over their respective alleles. When these plants were self-pollinated, 556 seeds were obtained, of which 315 were round and yellow, 101 wrinkled and yellow, 108 round and green, and 32 wrinkled and green. These are approximately in the proportions 9/16 round yellow, 3/16 wrinkled yellow, 3/16 round green, and 1/16 wrinkled green, a ratio of 9:3:3:1. Other similar crosses yielded these same ratios.

From these data Mendel formulated his second law, or the **Law of Independence**. The principle is a very simple one: it states that the members of one pair of genes segregate independently of other pairs. From this hypothesis he could deduce the results of other crosses and test for agreement with the data actually obtained.

If the characters are behaving independently, it means that the results of an experiment involving several sets of characters may be obtained by mutiplying together the proportions expected for each factor considered by itself. This may be illustrated for the data just mentioned. Letting *R* stand for round, *r* for wrinkled, *Y* for yellow, and *y* for green, the cross may be diagrammed as follows:

Considering shape only:



expected if the character were completely dominant. A conspicuous example is the coat color in Shorthorn cattle. The genotypes and phenotypes are as follows:

Genotypes	Phenotype
R/R	Solid color (Red in Shorthorns)
R/r	Mottled (Roan)
r/r	White

The probability of the occurrence of various kinds of offspring of two roan parents, as determined by the usual method, would be: red, 1/4; white, 1/4; roan, 1/2. In a very large family these expected proportions would be approximately realized.

FURTHER EXPERIMENTS OF MENDEL (Mendel's Second Law)

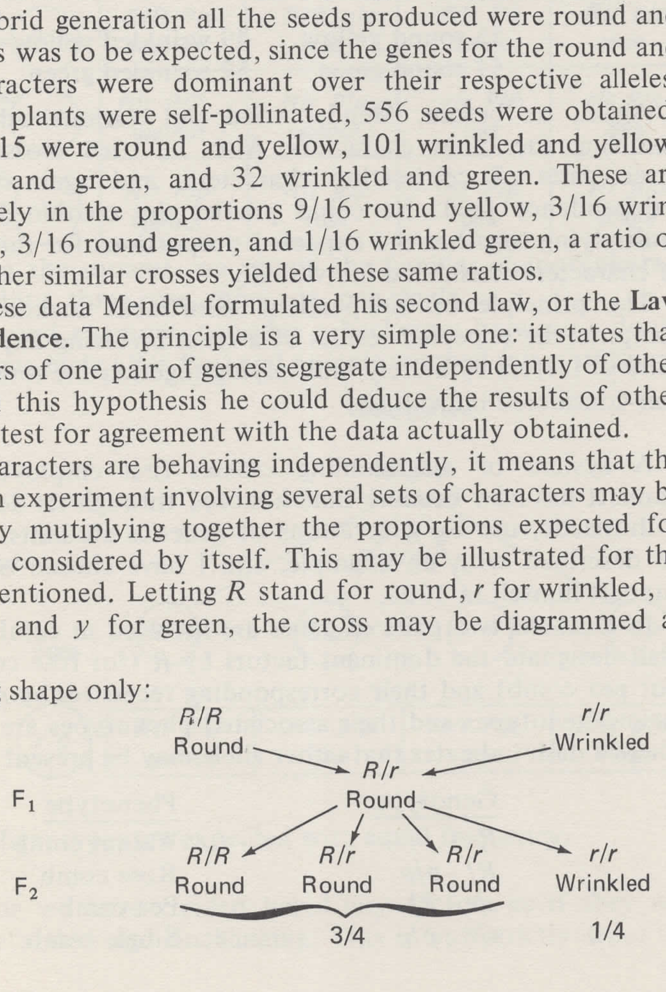
Mendel's first law, the law of segregation, applies to the behavior of a single pair of factors. The second law describes the simultaneous behavior of two or more pairs of genes.

Mendel made several hybridizations in which two or three pairs of factors were studied simultaneously. For example, he crossed plants producing round yellow seeds with plants producing wrinkled green seeds and observed the F_1, F_2, and F_3 generations. In the first hybrid generation all the seeds produced were round and yellow. This was to be expected, since the genes for the round and yellow characters were dominant over their respective alleles. When these plants were self-pollinated, 556 seeds were obtained, of which 315 were round and yellow, 101 wrinkled and yellow, 108 round and green, and 32 wrinkled and green. These are approximately in the proportions 9/16 round yellow, 3/16 wrinkled yellow, 3/16 round green, and 1/16 wrinkled green, a ratio of 9:3:3:1. Other similar crosses yielded these same ratios.

From these data Mendel formulated his second law, or the **Law of Independence**. The principle is a very simple one: it states that the members of one pair of genes segregate independently of other pairs. From this hypothesis he could deduce the results of other crosses and test for agreement with the data actually obtained.

If the characters are behaving independently, it means that the results of an experiment involving several sets of characters may be obtained by mutiplying together the proportions expected for each factor considered by itself. This may be illustrated for the data just mentioned. Letting R stand for round, r for wrinkled, Y for yellow, and y for green, the cross may be diagrammed as follows:

Considering shape only:



Considering color only, in exactly the same manner, the expectation in the F_2 generation is 3/4 yellow and 1/4 green. Multiplying these,

$$(3/4 \text{ Round} + 1/4 \text{ Wrinkled}) \times (3/4 \text{ Yellow} + 1/4 \text{ Green})$$

9/16 Round Yellow + 3/16 Round Green + 3/16 Wrinkled Yellow + 1/16 Wrinkled Green.

Mendel checked this hypothesis by predicting the offspring in other crosses. For example, he crossed the F_1 hybrid with the wrinkled green parent. Considering the shape only, he would expect 1/2 round and 1/2 wrinkled. Considering color only, he would expect 1/2 yellow and 1/2 green. Combined, according to the hypothesis of independence, he would expect

(1/2 yellow + 1/2 green) x (1/2 round + 1/2 wrinkled) = 1/4 round yellow + 1/4 round green + 1/4 wrinkled yellow + 1/4 wrinkled green.

The actual numbers obtained were

55 round yellow	49 wrinkled yellow
51 round green	52 wrinkled green

Again, as before, Mendel tried out various other matings involving the other characters. Also, he tried crosses involving three pairs of contrasting characters, and again the rule of independence held. He could predict the combined results by multiplying together the expected proportions for the three pairs of characters considered individually.

The principle of independent assortment makes it easy to compute the expected results of a cross involving any number of pairs of factors by merely multiplying together the results for each pair considered individually.

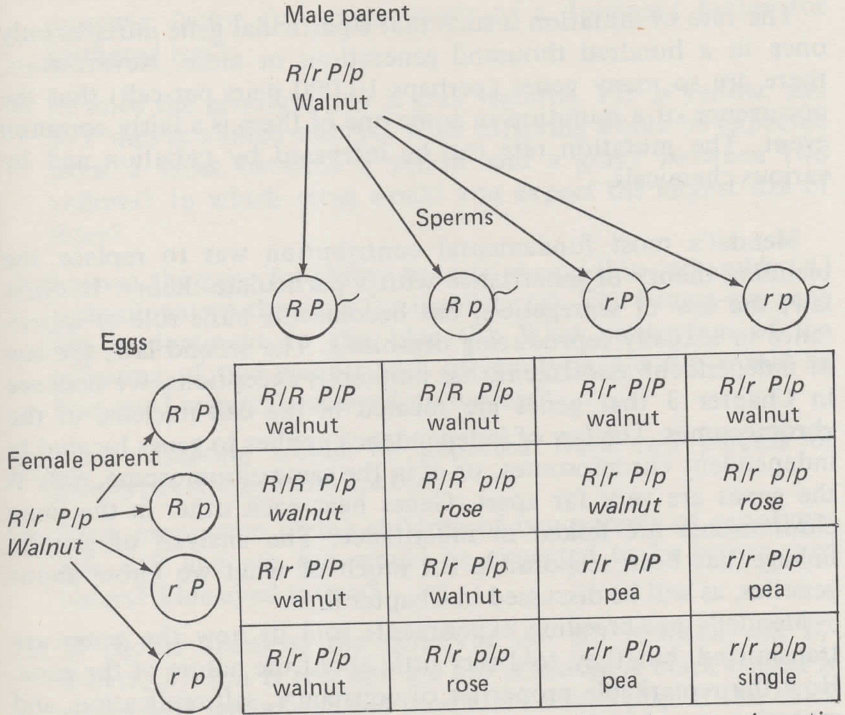
THE PUNNETT SQUARE

A system of diagramming crosses first employed by R.C. Punnett has now become time-honored through its dogged, if not enthusiastic, use by generations of genetics students. It can best be described with an example, and I have chosen one used by Punnett himself in 1905.

In chickens two pairs of genes are involved in comb shape. We shall designate the dominant factors by *R* (for rose comb) and *P* (for pea comb) and their corresponding recessives by *r* and *p*. The various genotypes and their associated phenotypes are as follows, where a dash indicates that either allele may be present.

Genotype	Phenotype
R/ - *P*/ -	Walnut comb
R/ - *p*/*p*	Rose comb
r/*r* - *P*/ -	Pea comb
r/*r* *p*/*p*	Single comb

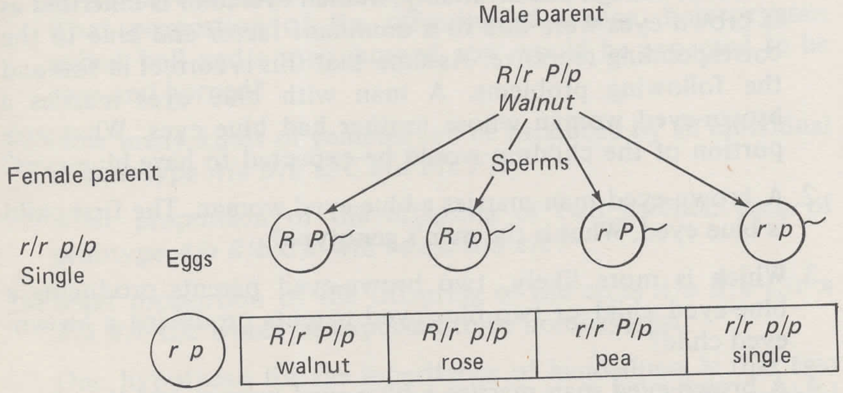
Consider first a cross between two birds, each heterozygous for both factors (i.e., of genotype *R*/*r* *P*/*p*). Since each gamete produced has one, and only one, representative of each pair of genes there are four possible kinds from each parent: *RP*, *Rp*, *rP*, and *rp*. These can be combined as shown in the square below.


Male parent
R/r P/p
Walnut
Sperms
Eggs
Female parent
R/r P/p
Walnut

	R P	R p	r P	r p
R P	R/R P/P walnut	R/R P/p walnut	R/r P/P walnut	R/r P/p walnut
R p	R/R P/p walnut	R/R p/p rose	R/r P/p walnut	R/r p/p rose
r P	R/r P/P walnut	R/r P/p walnut	r/r P/P pea	r/r P/p pea
r p	R/r P/p walnut	R/r p/p rose	r/r P/p pea	r/r p/p single

Counting the squares, we obtain the familiar phenotypic ratio of nine walnut, three rose, three pea, one single.

As a second example, consider a backcross between a doubly heterozygous male (*R*/*r* *P*/*p*) and a single combed (*r*/*r* *p*/*p*) female.


Male parent
R/r P/p
Walnut
Sperms
Female parent
r/r p/p
Single
Eggs

	R P	R p	r P	r p
r p	R/r P/p walnut	R/r p/p rose	r/r P/p pea	r/r p/p single

All four phenotypes are expected with equal frequency.

MUTATION

Thus far we have treated hereditary factors as if they were absolutely stable and unchanging. This is not strictly true; if it were there would be no evolution. On rare occasions a gene does change to another allele. For example, a gene producing yellow peas may change to one producing green. Such a rare change in a gene is called a **mutation** and the changed gene is called a **mutant**. A mutant gene is just as stable as it was in the original state, and is inherited in the same way.

The rate of mutation is such that a particular gene mutates only once in a hundred thousand generations or more. Nevertheless, there are so many genes (perhaps 10,000 pairs per cell) that the occurrence of a mutation in some one of them is a fairly common event. The mutation rate can be increased by radiation and by various chemicals.

HOW WELL HAS MENDEL'S WORK STOOD UP?

Mendel's most fundamental contribution was to replace the blending theory of inheritance with a particulate theory. His first law, the law of segregation, has become the basic rule of inheritance in sexually reproducing organisms. The second law, the law of independent assortment, has numerous exceptions. We shall see in Chapter 3 that genes are located in the cell nucleus, in the **chromosomes**. The law of independence applies to genes located in independent chromosomes, or, if in the same chromosome, only if the genes are very far apart. Genes near each other in the same chromosome are **linked** in inheritance. The analysis of genetic linkage has been responsible for much of what we know about genetics, as will be discussed in Chapter 8.

Mendel's pea-breeding experiments told us how the genes are transmitted, but they told very little about the nature of the gene. How the remarkable properties of constancy, self-replication, and mutation can be accounted for by the discoveries of molecular biology is the subject of the next chapter.

PROBLEMS

1. Usually, though not invariably, human eye color is inherited as if brown eyes were due to a dominant factor and blue to the corresponding recessive. Assume that this is correct in this and the following problems. A man with blue eyes marries a brown-eyed woman whose mother had blue eyes. What proportion of the children would be expected to have blue eyes?
2. A brown-eyed man marries a blue-eyed woman. The first child is blue eyed. What is the man's genotype?
3. Which is more likely, two brown-eyed parents producing a blue-eyed child or two blue-eyed parents producing a brown-eyed child?
4. A brown-eyed man marries a blue-eyed woman and they have eight children, all with brown eyes. Can you be sure whether the man is homozygous or heterozygous? Which is more likely? If the ninth child has blue eyes, will that settle the question?

5. What offspring would be expected from mating a roan cow with a roan bull? A roan with a white? A red with a white?

6. What would be the easiest way to determine whether a rose combed rooster is homozygous or heterozygous?

7. Which would be easier to get rid of in a flock of chickens, a recessive factor for single comb or a dominant factor for feathered legs?

8. In mice the genotype *y*/*y* is gray (agouti), *Y*/*y* is yellow, and *Y*/*Y* dies as a small embryo. What offspring would be expected from a cross between a yellow and a gray? Between two yellows? In which cross would you expect the largest size of litter?

9. In man the gene for ability to taste phenyl-thio-carbamide (*A*) is dominant to the gene for inability (*a*). The brown-eye gene (*B*) is dominant to the blue (*b*). What proportion of the offspring of two parents each of genotype *A*/*a* *B*/*b* would be blue-eyed tasters? blue-eyed nontasters?

10. What offspring would be expected from two parents of genotypes *A*/*a* *B*/*b* and *a*/*a* *b*/*b*?

11. What proportion of the children of two parents of genotypes *A*/*a* *B*/*B* and *a*/*a* *b*/*b* would be expected to be brown-eyed tasters? Blue-eyed tasters?

12. In cocker spaniels, the genotype *A*/- *B*/- is black, *a*/*a* *B*/- is liver, *A*/- *b*/*b* is red, and *a*/*a* *b*/*b* is lemon. A black cocker is mated to a lemon one and produces a lemon pup. If this black dog is mated to another of his own genotype what proportions would be expected among the offspring?

13. In cattle, polled is dominant to horned, and roan is the result of the heterozygous condition of the genes for red and white. What proportion of the offspring of a roan, heterozygous polled bull and a roan, horned cow would be expected to be roan and horned?

14. How many kinds of gametes can be produced by an individual of genotype *A*/*a* *B*/*b* *C*/*C* *d*/*d* *E*/*e* *F*/*F*?

15. What proportion of the offspring of two parents, each of genotype *A*/*a* *B*/*b* *C*/*c* will be *a*/*a* *b*/*b* *c*/*c*?

16. What proportion of the offspring of the cross *A*/*a* *B*/*b* *C*/*c* x *A*/*a* *B*/*b* *C*/*c* would be expected to be homozygous?

17. One hypothesis for the inheritance of handedness is that two loci are involved. At the first locus *L*/- produces left cerebral dominance and *l*/*l* right. At the second *C*/– causes the dominant hand to be contralateral (opposite) to the side of the dominant cerebral hemisphere and *c*/*c* causes the hand and brain dominance to be ipsolateral (on the same side). On

this hypothesis which of the following would be left-handed: *C*/- *L*/-, *C*/- *l*/*l*, *c*/*c* *L*/-, and *c*/*c* *l*/*l*?

18. Give an example of genotypes such that two left-handed parents would produce only right-handed children. Only left-handed. Both.

2

THE CHEMICAL NATURE OF THE GENE

WHAT IS A GENE?

The great British geneticist and statistician, R. A. Fisher, once said that the gene can be regarded from either of two viewpoints: (1) as a hypothetical entity by which we can make sense of breeding experiments, or (2) as a very tiny piece of meat. The first corresponds to the Mendelian view. Mendel's experiments showed that the principles of heredity can be formulated as very simple rules based on the behavior of particles which are transmitted from cell to cell and from parent to offspring. The nature of the particles, and even whether such particles exist, was unknown.

The history of genetics from the rediscovery of Mendel's laws in 1900 until the present time has been the gradual replacement of the Mendelian view by Fisher's second view. By 1950 there was strong evidence that the gene is **DNA** (for **Deoxyribonucleic Acid**) and, thanks to the happy insight of James Watson and Francis Crick, the chemical structure of DNA has been known since 1953. The "tiny piece of meat" that Fisher talked about is a segment of DNA.

For many purposes, such as animal and plant breeding and the study of human pedigrees, it is sufficient to regard the gene simply as an unanalyzed particle, and to be concerned solely with the rules of transmission. But, we are not content with this, and the most exciting recent advances in genetics are concerned with the nature and behavior of the gene itself.

WHAT THE GENE MUST BE ABLE TO DO

At the time of Mendel's experiments it was known that all organisms are composed of cells and that all the cells in the body are derived from a single cell. It was also known that the sperm (or pollen) and egg form the bridge between generations. From this knowledge and from the rules of inheritance and mutation, we can infer that the gene must be able to do at least four things. These are:

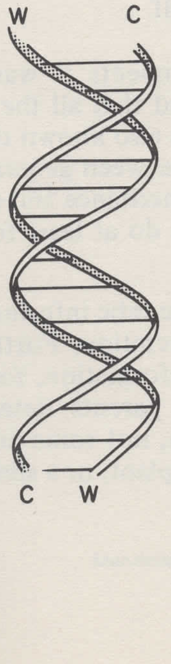
1. The gene must store and carry genetic information from cell to cell and from generation to generation. Furthermore, it must carry an enormous amount of information, for the set of genes that you inherited from your parents determined that the fertilized egg developed into you, not some other person, or an elephant, or a dinosaur, or an eggplant, or a slime mold.

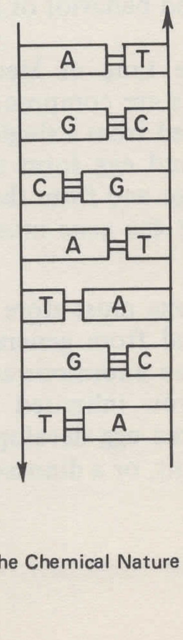
2. It must copy itself every time the cell divides. Moreover, it must copy itself with great precision, for the error rate is typically less than one per million copyings.
3. When it does make an error, the error must be copied in each successive cell division. In other words, there is mutation. This is a very unusual property in nonliving systems, but of vital importance for life, for without mutation there could be no natural selection as we understand it.
4. The information stored in the gene must be such that it can be decoded and translated into action. Somehow, the gene influences development to produce the character; or, in genetic parlance, the genotype (together with the environment) produces the phenotype.

We shall soon see that the structure of DNA immediately suggests mechanisms for the first three properties, and points the way to understanding the fourth.

THE STRUCTURE OF DNA

The double-helical structure of DNA, proposed by Watson and Crick, has turned out to be correct in almost every detail. The molecule is a very long, thin, twisted, double thread. The backbone of the molecule is a chain of alternating sugar and phosphate groups. The sugar is a 5-carbon type, **deoxyribose**. There are two of these chains, bridged at each sugar by a pair of **bases**. These are of four kinds: two large **purines (adenine** and **guanine**), and two small **pyrimidines (thymine** and **cytosine**). I will abbreviate them **A**, **G**, **T**, and **C**. Between the two sugar-phosphate chains there is room for one purine and one pyrimidine, but the space is too large for two pyrimidines and too small for two purines. Furthermore, because of their chemical structures, A and T can be held together by weak chemical bonds (hydrogen bonds) and so can G and C, but not other combinations. The structure is like a stepladder in which the sugar-phosphate chains are the uprights and the pairs of





bases, either AT or GC, are the steps. The whole structure is coiled so that it makes a complete rotation every 10 steps.

The diagram on the left shows the extent of the coiling of the uprights and the distance between the bridging steps. The steps, like those in a real stepladder, are flat. The diagram at the right shows the molecule as if it were untwisted (thereby sweeping under the rug a number of topological complexities). As shown, there is room between the uprights for one purine and one pyrimidine, and the hydrogen bond pairing rules (two bonds between A and T, three between C and G) determine the four possible kinds of steps.

There is a polarity to the molecule, arising from the asymmetrical three-dimensional structure of the sugar. I will omit the details until a later chapter, but the consequence is that in the diagram, the W strand viewed from the top looks exactly like the C strand viewed from the bottom.

A feature of a great scientific model is that it explains a number of otherwise unaccountable facts. The Watson-Crick structure of DNA is a beautiful example, for once we see the structure we immediately think of plausible explanations for the four gene properties listed above. We shall now consider these in turn.

DNA AS A CARRIER OF INFORMATION

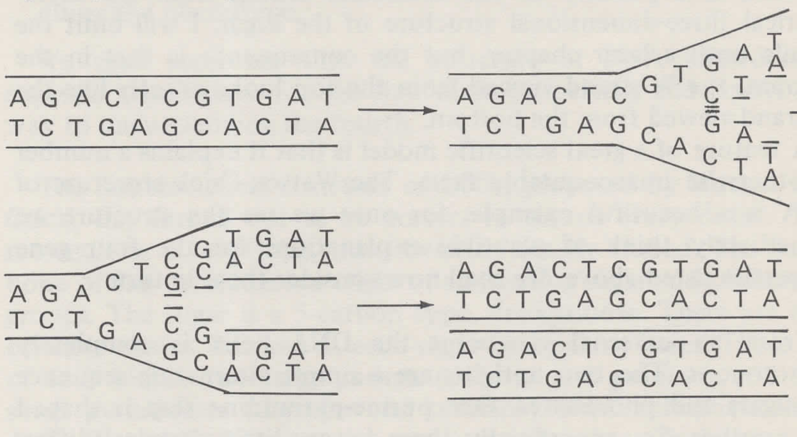
From an external viewpoint the DNA helix is completely monotonous. The two uprights are a simple alternating sequence of sugars and phosphates. One purine-pyrimidine step is shaped like another. So, superficially there is very little chemical difference between one gene and another. On the other hand, the stair-steps are of four kinds and there is no chemical reason for them to be in any fixed order; the molecule has the same helical shape regardless of the order.

This immediately suggests that the difference between one gene and another is the sequence of bases. This is how the gene stores and transmits information: the genetic message is spelled by the sequence of steps. Since there are four possibilities for each step, AT, TA, CG, and GC, a gene that is 1000 steps long could carry any of 4^{1000} different messages. This is an enormous number, far more than the number of elementary particles in the universe. So we have no trouble conceiving of the genes as capable of carrying an enormous amount of information encoded in the sequence of bases in the DNA molecule. We understand the first property.

Notice that, once we have specified the bases in one half of the double helix, we know the other half; this being fixed by the rule that opposite each T is an A, and opposite each G, a C. So half of the molecule is redundant. From the standpoint of information-carrying, there is no obvious reason for this. But from the standpoint of replication there is, as the next section will show.

DNA REPLICATION

The second property of the gene, that of self-copying, is immediately suggested by the complementary nature of the two strands of the helix. The two strands are easily separated, for the hydrogen bonds that hold together the two parts of a step are very weak compared to the other chemical bonds in the molecule. When the two strands separate, each of them has a structure that fits the component parts of the opposite strand. So opposite a T an A would fit naturally, opposite a G a C would fit, and so on. A, G, T, and C joined to the appropriate sections of the upright (this combination of a base, sugar and phosphate is called a **nucleotide**) are present in the nucleus of the cell. So, the two strands come apart somewhat like a zipper; each part picks up nucleotides complementary to itself, and two new complementary strands are built up. The nucleotides of the new strand are fastened together by enzymes. The process can be visualized roughly in this way:



The redundancy of the double helix, in having one strand that adds nothing to the information carried by the other, now makes sense. The double structure is required for the opposite-copying mechanism of replication. More details are in Chapter 12.

THE MOLECULAR BASIS OF MUTATION

The third property to be explained is mutation; that is to say, how an error in the gene can be replicated. The DNA structure immediately suggests the answer. Any change in one or more of the bases, for any reason, will dictate that the complement of the new arrangement will be synthesized rather than that of the old. The copying mechanism acts on whatever is there at the moment, not what was formerly there.

We would expect random changes for all sorts of reasons. They should (and do) happen spontaneously, since molecules make mistakes. We would also expect that processes that induce chemical reactions (such as radiation) or chemicals that act directly on DNA would enhance the rate of errors; then the replicating mechanism dutifully copies these errors.

The basic mystery of mutation is solved: it is nothing more than random error and the blindly faithful copying mechanism.

TRANSLATING THE MESSAGE INTO ACTION

It is clear that the DNA molecule is capable of storing and transmitting an enormous amount of information from generation to generation. The information is in the form of a linear message, encoded in the sequence of nucleotides. It is also clear that there has to be some mechanism to translate this into control over cell reactions, into embryonic development, and into all the other activities that are gene-determined.

Proteins are the key to most of the important processes of growth and development. A protein, like DNA, is a long threadlike molecule, although it is folded into a complex three-dimensional structure on which its biological specificity depends. So the basic problem in converting the genetic message into action is translating a genetic message carried in the base-sequence of DNA into a sequence of amino acids that distinguishes this protein from another. This involves *transcribing* the DNA message into another type of nucleic acid, **ribonucleic acid**, or **RNA**, which in turn is *translated* into a sequence of amino acids in the process of protein synthesis. The nature of the code and the transcribing mechanisms are now well understood, but we shall defer discussion until later in the book (Chapter 12).

Phenotypic variety comes mainly from proteins. What makes organisms differ, or gives a particular individual its uniqueness, is the protein composition. We can say that the essence of the phenotype is the array of kinds, abundance, and organization of different proteins. On the other hand, we have just learned that the essence of the genotype is nucleic acid. To understand phenotypes, we must understand proteins; to understand genotypes, we must understand DNA; and to understand genotypic control of the phenotype, we must understand how the DNA determines the proteins. We'll return to this in Chapter 12.

EVIDENCE FOR THESE CONCLUSIONS

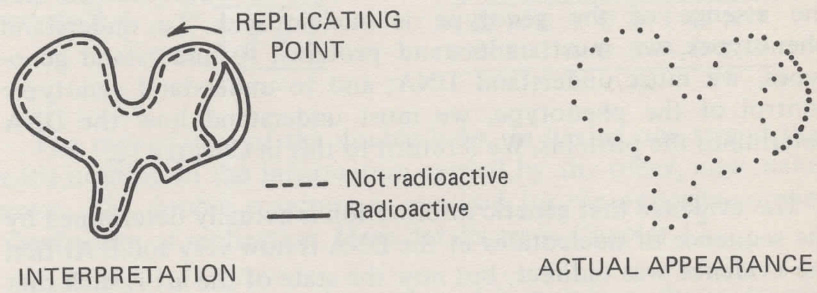
The evidence that genetic information is actually determined by the sequence of nucleotides in the DNA is now very solid. At first the evidence was indirect, but now the state of the art is such that the actual sequence of some of the smaller genes is known by direct chemical analysis. Recently H. Khorana achieved the actual synthesis of a gene. He succeeded in making in the laboratory a sequence of DNA identical to that of a known yeast gene.

Likewise, the genetic nature of many mutations is known from genetic and chemical analysis. In some cases the change is a replacement of one base by another. In others an extra base is inserted, or a base is deleted. In still others there are changes involving the gain, loss, or rearrangement of several bases.

The Watson-Crick rules of replication received their first strong experimental support from experiments of Meselson, Stahl, and Vinograd. Their experimental procedure made use of a density gradient in a centrifuge tube. A solution of cesium chloride in a high-speed centrifuge sets up a gradient of concentrations depending on the equilibrium between the centrifugal force and the random motion of the molecules. Within this continuous gradient DNA's of different density will come to rest at different distances from the center of rotation. The researchers grew bacteria for many generations in a medium containing heavy nitrogen (N^{15})

until virtually all the nitrogen in the cells was N^{15}. Then the cells were grown in a medium of N^{14} long enough to divide once. At that time the DNA was extracted from the cells and put into the density gradient centrifuge. The equilibrium position was exactly half way between that of N^{14} and N^{15}; thus these were hybrid molecules as would be expected if each double helix contained one old N^{15} and one new N^{14} strand. Furthermore, when the cells were permitted to grow two generations in N^{14} there were two kinds of molecules, one like the hybrid molecules and one like the N^{14}. Thus the predictions of the Watson-Crick model are verified.

Additional evidence comes from DNA labeled with radioactive tritium (in place of normal hydrogen). If labeled thymine is introduced at the beginning of DNA replication, then the new strand will be radioactive. After one replication, the double helix will have one radioactive and one nonradioactive strand. After the beginning of a second replication, the chromosome will look like the figure below, where the dashed line represents the old strand, and the new radioactive strands are represented by a solid line. (I have not mentioned previously that some DNA is circular; that is the case for the bacterial DNA used in these experiments.)



The chromosome is extracted from the bacterial cell, (very gently to avoid breaking the DNA), and placed on a photographic emulsion. The electrons emitted by the radioactive tritium strike the emulsion close to the chromosome, leading to a spot on the photograph, and the DNA thus takes its own picture. This work was first done by J. Cairns, and in his photographs you can easily see that there are twice as many dots per unit length in regions, such as in the right of the diagram, where both strands of the DNA are radioactive.

It is rare in the history of science, and very rare in biology, for a simple idea to appear and suddenly make the most baffling processes understandable. Prior to the Watson-Crick model, hypotheses for information storage, copying, and mutation were purely speculative and none was satisfying or had any experimental support. The DNA model immediately suggested ways in which it could be tested. The further merit is that it also suggested a host of further testable hypotheses, and molecular biology was off and running.

PROBLEMS

1. How could you tell from base composition whether a DNA were single or double stranded?
2. In the DNA sequence $\begin{matrix}\text{ACTTTCCATGTTAT}\\ \text{TGAAAGGTACAATA}\end{matrix}$ the strand is broken after the third base from the left and before the third base from the right. Write the new sequence after the inner piece has inverted and reattached.
3. How many different messages can be carried by a DNA double helix 100 bases long? How many messages could it carry if it were restricted to A and T?
4. If three bases are required to code for one amino acid, how long must the gene be to encode a protein containing 400 amino acids?
5. A chromosome is labeled with radioactive thymine and then allowed to divide twice in cold medium. How many of the four chromosomes would be radioactive?
6. The adenine molecule can, with very low probability, momentarily shift one hydrogen so that it pairs with cytosine. Explain how this might cause mutation.
7. The molecular weight of the DNA of a bacterial nucleus is about 3×10^9. The molecular weight of a single nucleotide pair is about 660. How many nucleotide pairs are there in the chromosome? If each nucleotide occupies a length of 3.4 angstroms (1 angstrom = 10^{-7} mm), how many millimeters long is the bacterial DNA?
8. A haploid mammalian cell has about 3×10^9 nucleotide pairs. How long would the DNA of an average cell be if put end to end? (If you like this sort of thing, how many miles would all the DNA in the human body extend? Assume about 10^{14} cells. Would it reach the moon? Distance from earth to the moon: 238,857 miles.)
9. How many permutations of the four base-pair steps in the DNA helix can be made using three at a time? (In other words, how many three-letter words can be made with a four-letter alphabet?)
10. If there were only two bases, A and T, how many steps would be required to produce the same number of sequences?

CHROMOSOME BEHAVIOR AND THE CHROMOSOMAL BASIS OF HEREDITY

3

THE DISCOVERY OF CELLS

Soon after the invention of the microscope many curious people began to study animal and plant tissues. Such study led to the observation that tissues are composed generally of discrete units which were named cells. By the latter half of the nineteenth century microscopes of greater resolving power had been developed and superior methods of preparation of tissues for microscopic study were known so that it became possible to study the internal contents of cells in some detail.

STRUCTURE, SIZE AND NUMBER OF CELLS

A typical cell contains an inner structure known as the **nucleus** while the remaining material is known as the **cytoplasm.** Within the nucleus there are a number of threadlike or worm-shaped bodies known as **chromosomes.** There are also a number of structures that are sometimes found in the cytoplasm (**mitochondria, plastids, ribosomes,** etc.), most of which have had very little emphasis in classical genetics.

The size of cells ranges between wide extremes. Some nerve cells are several feet long in an elephant or a giraffe. A hen's egg or ostrich egg may be regarded as a single cell. At the other extreme are tiny bacterial cells which are in the vicinity of one micron (1/1000 of a millimeter or 1/25,000 inch) in diameter. Most of the cells which make up the bodies of plants and animals are around 10 microns in diameter. An adult human has perhaps 250,000,000,000,000 cells in his body. If all cells were dividing synchronously (which is by no means the case), it would require some 50 generations of cell divisions to produce this number starting with a single fertilized egg.

Each species has a characteristic chromosome number. For example, one species of the intestinal round worm, Ascaris, has only two chromosomes per cell. Drosophila has 8, the mouse has 40, and the common grasshopper has 24 in the female and 23 in the male. The number of chromosomes is no index of complexity, however, for some invertebrates have several hundred. Among plants the numbers also vary tremendously; the hawksbeard (Crepis) has 6 per cell, Mendel's peas had 14, corn has 20, the common potato has 48, but there are other plants whose number exceeds 200.

The human chromosome number is 46. Whereas the number of chromosomes in cells of different tissues is remarkably constant, cell cultures grown *in vitro* and malignant tumors may have highly variable numbers.

The chromosomes occur in pairs and, with the exception of the sex-chromosomes to be discussed in the next chapter, the two members of a pair are indistinguishable. However, the members of different pairs can often be distinguished by size and structural differences. In highly studied species the chromosome pairs are given designations, usually a number.

MITOSIS (Somatic Nuclear Division)

When a cell divides, the cytoplasmic constituents pass more or less at random to one or the other of the two cells. The chromosomes, on the other hand, go through a very precise process which insures that the same number of chromosomes gets into each of the new cells. Cell division, although it is a continuous process, is conventionally described in five phases according to the arrangement of the chromosomes.

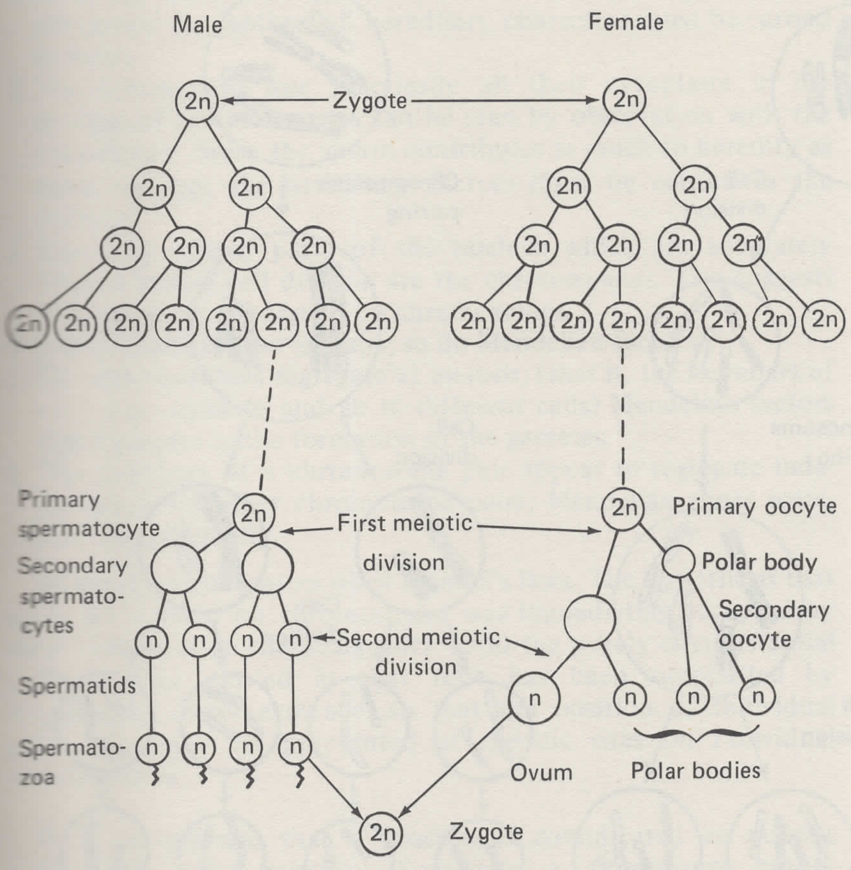
1. Interphase. The chromosomes are stretched out in the form of long threads and are practically invisible with a microscope. During this phase each chromosome is duplicated. Each of the two parts is called a **chromatid.**
2. Prophase. The chromosomes begin a process of shortening and thickening, which appears to be due to a coiling of the chromosome thread.
3. Metaphase. By this stage, the chromosomes have shortened until they are only a small fraction of their former length, but are much thicker. The chromosomes are now lined up in a single plane across the center of the cell.
4. Anaphase. The two members of each chromosome pair start to be pulled apart by a threadlike spindle fiber that attaches to a specific place, the **centromere**, on the chromosome.
5. Telophase. The entire cell is now almost completely divided and the chromosomes in each part are beginning to uncoil and lengthen. The cells finally become completely divided and the chromosomes reach again the interphase condition.

The most important feature of mitosis from the standpoint of the study of heredity is that the chromosome number remains constant through successive cell divisions. There is a chromosome duplication followed by a cell division, another chromosome duplication, another cell division, and so on. The result is an exact distribution of chromosomes to the new cells as they are formed, and the chromosome makeup of a cell is the same as that of the cell from which it has been derived.

MEIOSIS (Formation of Gametes)

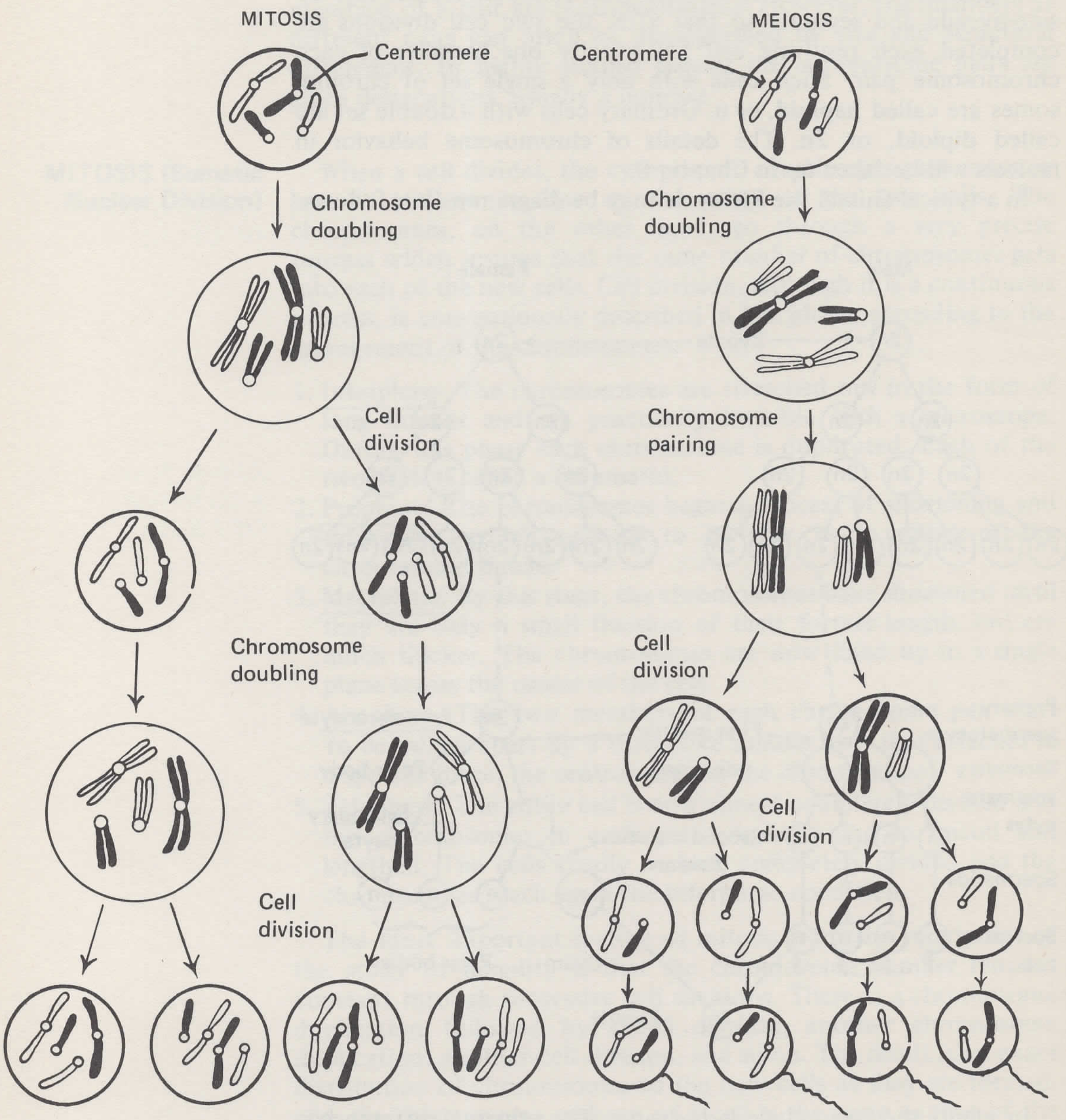
In ordinary mitosis there is a chromosome duplication, in which the number of chromosomes is doubled, followed by a cell division, in which the number in each cell is halved. This cycle continues for cell division after cell division, the total number of chromosomes per cell remaining the same. In the process of **meiosis**, just before the formation of the reproductive cells or gametes, there is a change in this regular process. There is a chromosome duplication followed by two cell divisions so the resulting cells have only half the usual number of chromosomes. Furthermore, the members of each pair of chromosomes line up side-by-side and separate so that after the two cell divisions are completed each resulting cell has exactly one member of each chromosome pair. Such cells with only a single set of chromosomes are called **haploid**, or **n**. Ordinary cells with a double set are called **diploid**, or **2n**. The details of chromosome behavior in meiosis will be taken up in Chapter 8.

In a typical animal the life cycle may be diagrammed as follows:


Male
Female
Zygote
2n
Primary spermatocyte
Secondary spermatocytes
First meiotic division
Second meiotic division
Primary oocyte
Polar body
Secondary oocyte
Spermatids
Spermatozoa
n
Ovum
Polar bodies
Zygote

In plants the basic cycle is the same. The principal difference is that the haploid cells may undergo several mitotic divisions before the fusion of cells to form the diploid zygote. In some lower plants the number of cell divisions in the haploid phase greatly exceeds that in the diploid, but in higher plants the situation is reversed. The haploid phase is known as the **gametophyte**, the diploid as the **sporophyte**.

COMPARISON OF MITOSIS AND MEIOSIS


MITOSIS
Centromere
Chromosome doubling
Cell division
Chromosome doubling
Cell division
MEIOSIS
Centromere
Chromosome doubling
Chromosome pairing
Cell division
Cell division

And so on. Each cell retains the normal chromosome number.

Each sperm cell has half the normal number.

The black chromosomes came from one parent, the white from the other.

THE CHROMOSOME THEORY OF HEREDITY

The first widely accepted statement of the rule that all animals and plants are made up of cells was made in 1838 by Schleiden and Schwann. The discoveries of sperm and egg cells and the fact that the egg would not develop without fertilization were made about 1850. Mendel's breeding experiments were reported in 1865, so he had a chance to be familiar with the cell theory and the importance of fertilization. Chromosomes were first described about 1875, and by the turn of the century the details of mitosis and meiosis were quite well known. Immediately after the discovery of Mendel's rules in 1900 several people pointed out the possibility that the hereditary factors of Mendel were carried on chromosomes. The most convincing case was made by Sutton in 1903, whose arguments were essentially as follows:

1. Since the sperm and egg cells provide the only bridge from one generation to another, all hereditary characters must be carried in them.
2. The sperm cells lose practically all their cytoplasm in the process of maturation, as can be seen by observation with the microscope. Since the sperm contributes as much to heredity as does the egg, the hereditary factors must be carried in the nucleus.
3. The only visible parts of the nucleus which are accurately divided during cell division are the chromosomes. This suggests that the genes are carried on chromosomes.
4. Chromosomes occur in pairs; so do Mendelian factors.
5. The chromosomes segregate at meiosis; (that is, the members of each pair separate and go to different cells) Mendelian factors also segregate in the formation of the gametes.
6. The members of a chromosome pair appear to segregate independently of other chromosome pairs; Mendelian genes segregate independently.

In short, chromosomes obey Mendel's laws. The hypothesis that genes are carried on chromosomes was immediately accepted by most biologists. In the years since 1900 the purely circumstantial evidence that existed at that time has been supplanted by increasingly direct evidence, so that the position of individual genes can now be pinpointed at specific sites on individual chromosomes.

CHROMOSOME LOSS AND NONDISJUNCTION

It is no surprise that a process as complicated as meiosis occasionally makes mistakes. Sometimes a chromosome simply *gets lost – when a nucleus divides into two*, a chromosome *gets* left out of both. If this happens in meiosis a gamete is produced that is lacking one chromosome, and the resulting zygote is missing a chromosome. Almost always this leads to some kind of observable abnormality, and often causes death. A complete set of chromosomes is necessary for normal development.

Sometimes a pair of chromosomes do not separate properly at anaphase with the result that both get included in one daughter nucleus. The result is one nucleus with too many chromosomes and the other with too few. When this happens at meiosis, gametes with too few or too many chromosomes are produced. The process is called **nondisjunction**.

One of the most common cases of severe mental retardation in human beings is **Down's syndrome**, often called **mongolism**. Children with this condition have a large number of physical abnormalities in addition to retardation. Since persons with Down's syndrome hardly ever reproduce, nothing was known about the inheritance for many years. But with the coming of modern techniques for studying human chromosomes it was discovered that this is caused by nondisjunction of chromosome number 21; persons with the condition have this chromosome represented three times instead of the normal two.

Two other syndromes in humans are known to be caused by extra chromosomes; one by chromosome number 13, the other by number 18. Both produce severe abnormalities and early death.

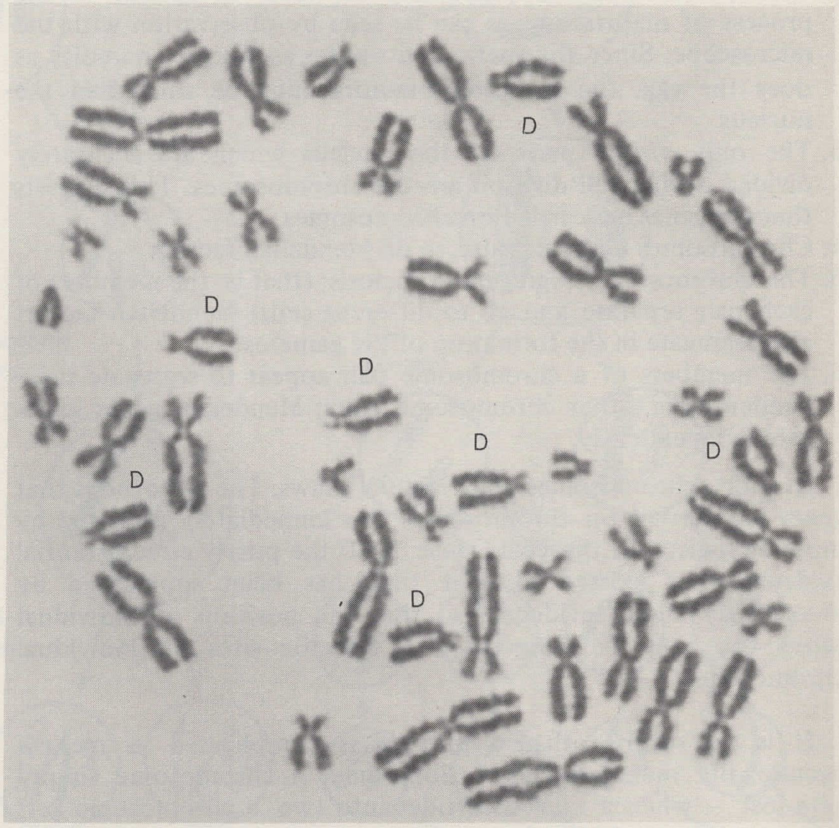
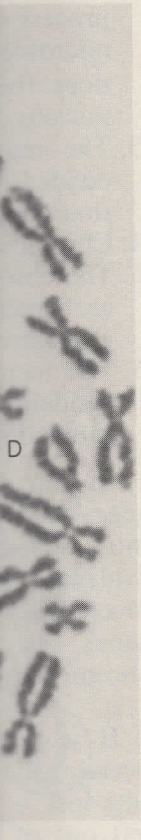


Figure 1. Human chromosomes. They are in the metaphase stage, just before division; the chromatids have separated except at the centromere. If you count these chromosomes, you will see that there are 47 instead of 46. This cell came from a boy who had an extra chromosome number 13. Since numbers 13, 14, and 15 are indistinguishable in this photograph, they are all labelled D; note that there are seven D chromosomes instead of the expected six.

omosomes
is, gametes
uced. The

ardation in
nongolism.
of physical
rsons with
vas known
coming of
nes it was
romosome
romosome

caused by
ne other by
arly death.



ge, just before
omere. If you
ad of 46. This
ber 13. Since
h, they are all
nstead of the

Numbers 13, 18, and 21 are all small chromosomes. All the other chromosomes lead to embryonic death when they are present in triplicate, as do all the chromosomes when one is missing. In general, a missing chromosome does more harm than an extra chromosome.

Curiously, and for reasons not now understood, meiotic nondisjunction is very much related to the age of the female. Women age 35 and above have a greater probability of having a child with an extra chromosome.

Finally, nondisjunction and chromosome loss are not confined to meiosis. They can also happen during mitosis. When this happens during embryonic development, the individual has two or more kinds of cells—a condition known as **mosaicism**. The loss of a chromosome responsible for some necessary part of chlorophyll production in a plant leaf can lead to a white spot.

IMPROVED METHODS FOR STUDYING CHROMOSOMES

The study of chromosomes has become much easier in recent years because of new technical improvements. This has been especially useful in the analysis of human chromosomes. One trick is to take advantage of the fact that the centromere is the last part of the chromosome to separate when the two chromatids pull apart, and catch the cell right at this time. (The drug, colchicine, arrests chromosome separation at this stage.) This enables the observer to determine easily the position of the centromere and this aids in identifying individual chromosomes with different centromere positions. Furthermore, new staining procedures have been found that color different parts of the chromosome differentially, so that each chromosome has a characteristic pattern. Using these procedures, a skilled cytologist can easily recognize each chromosome and knows to which of the 23 pairs it belongs. It is even possible to recognize individual parts of chromosomes if the chromosomes become broken or rearranged. Since chromosome aberrations often lead to severe physical abnormalities and mental impairment, it is of great medical importance to be able to identify individual chromosomes and study them in detail.

CHROMOSOMES AND DNA

In bacteria the chromosome is a simple DNA double-helix. It is easily observed in electron microscope photographs. The total length of the bacterial DNA, which is a single circular chromosome, is about 1 mm. Somehow this is stuffed into a cell whose diameter is only 1/1000 as large.

Organisms that have no distinct nucleus (such as bacteria and some of the more primitive algae), and whose chromosomes are simply DNA, are called **prokaryotes**. The higher plants and animals (**eukaryotes**) have a cell nucleus and a regular mitotic process.

The eukaryotic chromosome consists of both DNA and protein. The usual protein is **histone**, which has a high concentration of basic amino acids. This attaches to the acidic phosphate groups of the DNA, so there is a DNA-histone thread. In addition, there are a variety of nonbasic proteins whose function is not understood.

In the fruit fly, *Drosophila melanogaster,* from which we have learned so much about genetics, the amount of DNA in the largest chromosome contains about 65 million base pairs. Since the steps in the DNA ladder are 3.4×10^{-7} mm apart, this means that the total length of DNA in the chromosome is about 22 millimeters, or roughly one inch. Yet the chromosome as seen at metaphase is only a few thousandths of a millimeter in length.

Is the chromosome a single long DNA thread, or is it more like a cable with many strands? There is recent evidence for the former. The evidence comes partly from biophysical measurements based on the time required for the DNA to contract after it has been stretched—the longer the DNA, the longer the time required. Such measurements suggest that the DNA is very long, about the length expected if the chromosome consists of one long DNA molecule. Supporting evidence comes from electron micrographs of Drosophila DNA; again the length is consistent with there being a single DNA thread. How this thread is coiled into the chromosome, and how it keeps from becoming hopelessly tangled, remain mysteries.

The total length of DNA in the human cell is enormous. There are about 3.5×10^9 base pairs. This is the haploid amount, so the total length in a diploid cell is about 2 meters. With 23 pairs of chromosomes, the average length of DNA per chromosome is about 4 cm, or nearly 2 inches.

PROBLEMS

1. How many spermatozoa come from 100 spermatids? From 100 primary spermatocytes?
2. How many eggs come from 100 primary oocytes?
3. How many chromosomes are in a human zygote? Primary spermatocyte? Primary oocyte? Spermatid? Ovum? Spermatozoon?
4. If a character were determined by something carried in the cytoplasm would the offspring more often resemble the father or the mother?
5. How many kinds of gametes can an individual with 3 pairs of chromosomes produce? With 23 pairs?
6. How would you explain the occurrence of a person with one blue and one brown eye?
7. Very rarely a person with Down's syndrome has children. What would be the expected offspring from a woman with Down's syndrome who married a normal man?

4

SEX-CHROMOSOMES AND SEX-LINKAGE

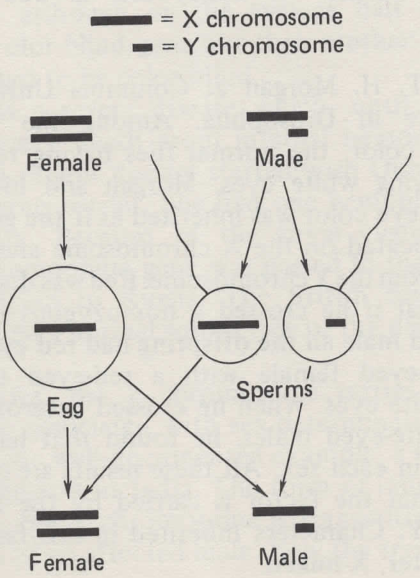
SEX-CHROMOSOMES

About 1900, when microscope techniques had become quite well developed and chromosome behavior was understood, it was noticed by several observers that there was one pair of chromosomes that differed from the others. The members of this pair were alike in one sex (usually the female) but different in appearance in the other sex. The two that were identical in one sex were the same as one of the members of the unlike pair in the other sex.

The chromosome which was present twice in the female and only once in the male was called the X chromosome. The odd chromosome was called the Y. Hence we can characterize the two sexes as follows:

XX = female
XY = male

A human male has an X chromosome, a Y chromosome and 22 pairs of other chromosomes, making a total of 46. The female has a pair of X's and 22 other pairs. The special pair are called **sex-chromosomes**. The others are **autosomes**.



SEX DETERMINATION

The sex-chromosomes segregate at meiosis just as the other chromosomes do, so this means that each sperm cell receives only one sex-chromosome. Hence in the process of spermatogenesis there will be two types of sperm cells produced in equal numbers, those containing an X chromosome and those containing a Y chromosome. Each of the eggs produced by a female will contain one X chromosome. Therefore, the sex of the future individual is determined at the time of fertilization of the egg. If the egg is fertilized by a sperm bearing a Y chromosome (along with 22 ordinary chromosomes), the zygote will have an X and a Y and will develop into a male. If the egg is fertilized by an X-bearing sperm, the zygote will have two X chromosomes and will develop into a female.

OTHER SYSTEMS OF SEX DETERMINATION

Most animals follow the same rule as Drosophila and man in having two X chromosomes in the female and an X and a Y in the male. The X and Y usually differ enough in shape and size to be recognizably different when viewed through a microscope, though this is not always true (e.g., in the mosquito they are identical, even after the most meticulous examination). Sometimes the Y chromosome is missing entirely, as in grasshoppers. In this case it is customary to speak of XX females and XO males.

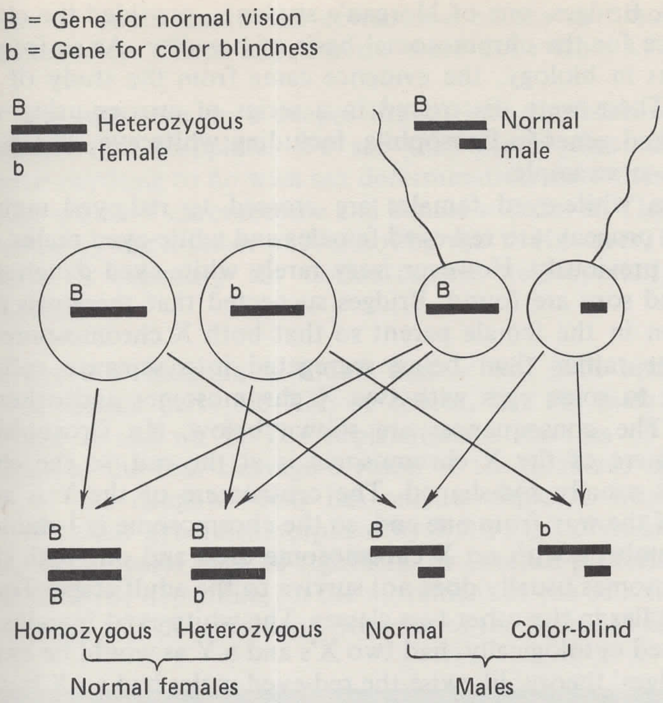
In the birds, moths, and butterflies (and in occasional species in other groups) the sex-chromosomes are reversed; the male has two X chromosomes and the female only one. The Y may or may not be present.

In the hymenoptera (bees, wasps, and ants), unfertilized eggs develop into males and fertilized eggs into females, so that the males are haploid and the females are diploid. The haploid male has a modified meiosis so that the sperms have a full haploid set of chromosomes. The development of various classes of females, such as queens and workers in honeybees, is due to differential nutrition.

SEX-LINKAGE

About 1910 T. H. Morgan at Columbia University began to study inheritance in Drosophila. Among the first characters studied was eye color, the normal flies having red eyes and the mutant type having white eyes. Morgan and his students soon noticed that the eye color was inherited as if the gene determining this trait were located on the X chromosome and this gene were entirely missing from the Y chromosome. Red was dominant to white.

He noticed that if he crossed a homozygous red-eyed female with a white-eyed male all the offspring had red eyes, but when he crossed a white-eyed female with a red-eyed male, the male offspring had white eyes. When he crossed heterozygous red-eyed females with white-eyed males, he found that half the offspring were white-eyed in each sex. All these results are compatible with the hypothesis that the factor is carried by the X chromosome, but not by the Y. Characters inherited in this fashion are called **sex-linked**, or better, **X-linked**.


B = Gene for normal vision
b = Gene for color blindness
B
b
Heterozygous female
B
Normal male
B
b
B
B
B
b
B
B
b
Homozygous
Heterozygous
Normal
Color-blind
Normal females
Males

The female, since she has two X chromosomes, may be either homozygous or heterozygous for an X-linked gene. The male, since he has only one representative of such a gene, is said to be **hemizygous.**

The best known examples of X-linked inheritance in the human are color blindness and hemophilia. An example of color-blind inheritance is shown in the diagram; in this example a heterozygous woman marries a man with normal vision. The daughters are all normal, although on the average half of them will be carriers of the color-blind gene like their mother. Of the sons, half would be expected to be color-blind.

Hemophilia is a severe disease which, until recently, almost always led to early death from internal hemorrhage. The most famous pedigree of this disease started with Queen Victoria, who was a heterozygous carrier. She had one hemophilic son and two carrier daughters. Because of the royal custom of exporting daughters, the hemophilia gene was passed to the royal families of Europe from Spain to Russia. The British royal family, being descended through a normal son, is free of the disease.

Y-LINKED INHERITANCE

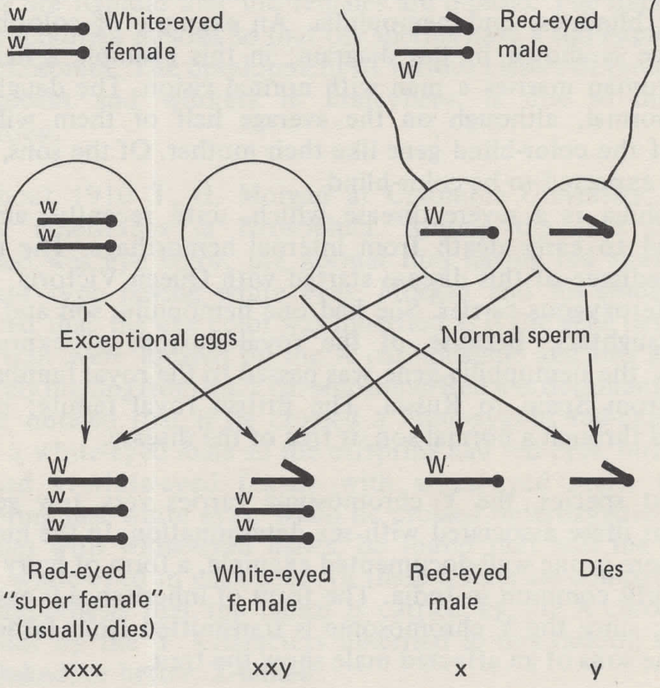
In most species, the Y chromosome carries very few genes, other than those associated with sex determination. In the human species there is one well-documented example, a form of hairy ears that is fairly common in India. The form of inheritance is easy to recognize; since the Y chromosome is transmitted from father to son, all the sons of an affected male show the trait.

NONDISJUNCTION AS PROOF OF THE CHROMOSOMAL BASIS OF HEREDITY

C. B. Bridges, one of Morgan's students, provided the clinching evidence for the chromosomal basis of heredity. As so frequently happens in biology, the evidence came from the study of exceptions. These were discovered in a series of crosses using various sex-linked genes in Drosophila, including white eye, which will be used as an example.

When white-eyed females are crossed to red-eyed males, the regular progeny are red-eyed females and white-eyed males, as was shown previously. However, very rarely white-eyed daughters and red-eyed sons are found. Bridges suspected that there was nondisjunction in the female parent so that both X chromosomes went together rather than being segregated into separate cells—thus leading to some eggs with two X chromosomes and others with none. The consequences are shown below. (In Drosophila the centromere of the X chromosome is at the end, so the chromosome is usually rod-shaped. The centromere of the Y is about a third of the way from one end, so the chromosome is J-shaped.)

An embryo with no X chromosome dies, and one with three X chromosomes usually does not survive to the adult stage. The main interest lies in the other two classes. The white-eyed females, when examined cytologically, had two X's and a Y as would be expected on Bridges' theory; likewise the red-eyed males had an X but no Y. Thus, every time there is an exception in the inheritance of the eye color there is a corresponding exception in the chromosome makeup—positive proof of the chromosomal basis of the white-eye gene (and of the other sex-linked genes that he studied). Historically, this experiment was of great importance for it removed the



last vestige of doubt as to the correctness of the chromosome theory of heredity. These experiments were done from 1913 to 1916.

These data also provide a deeper insight into the basis of sex determination in Drosophila. We see that the Y chromosome cannot have anything to do with sex determination since there can be a female with a Y chromosome and a male without it. Thus sex *is determined by the number of X's, not by the Y. However, the Y chromosome is necessary for normal sperm development; XO males* are sterile.

SEX CHROMOSOME MOSAICS (GYNANDROMORPHS)

Sometimes a Drosophila is found that is half male and half female. The female parts are XX, of course, but the male parts have a single X and no Y. The explanation is the loss of an X chromosome at the first cleavage division of the fertilized egg, so that one of the daughter cells has only a single X. Such sex mosaics are called **gynandromorphs**. Physically, the division into male and female parts may be right-left, or anterior-posterior, or more complicated, depending on the movement of the cells during the early embryonic development. Sometimes the fly is 3/4 female

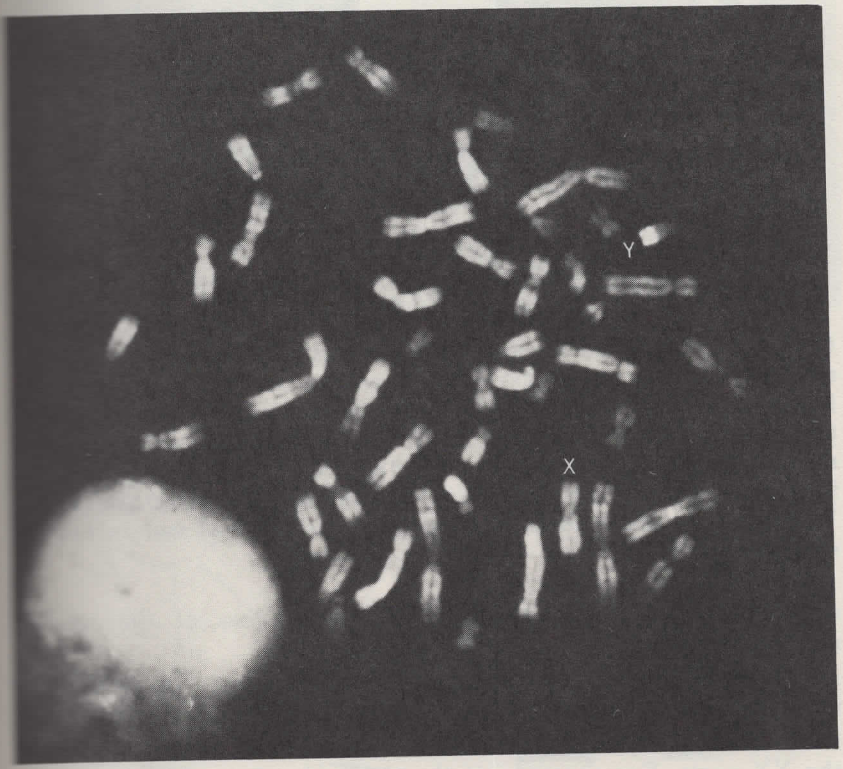


Figure 2. Human chromosomes stained with quinacrine, which fluoresces under ultraviolet light. The different chromosomes can be identified by characteristic fluorescence patterns (Q bands). The Y chromosome is especially bright and is easily recognized by this technique. (The large blob at the lower left is irrelevant: it is the nucleus of another cell.)

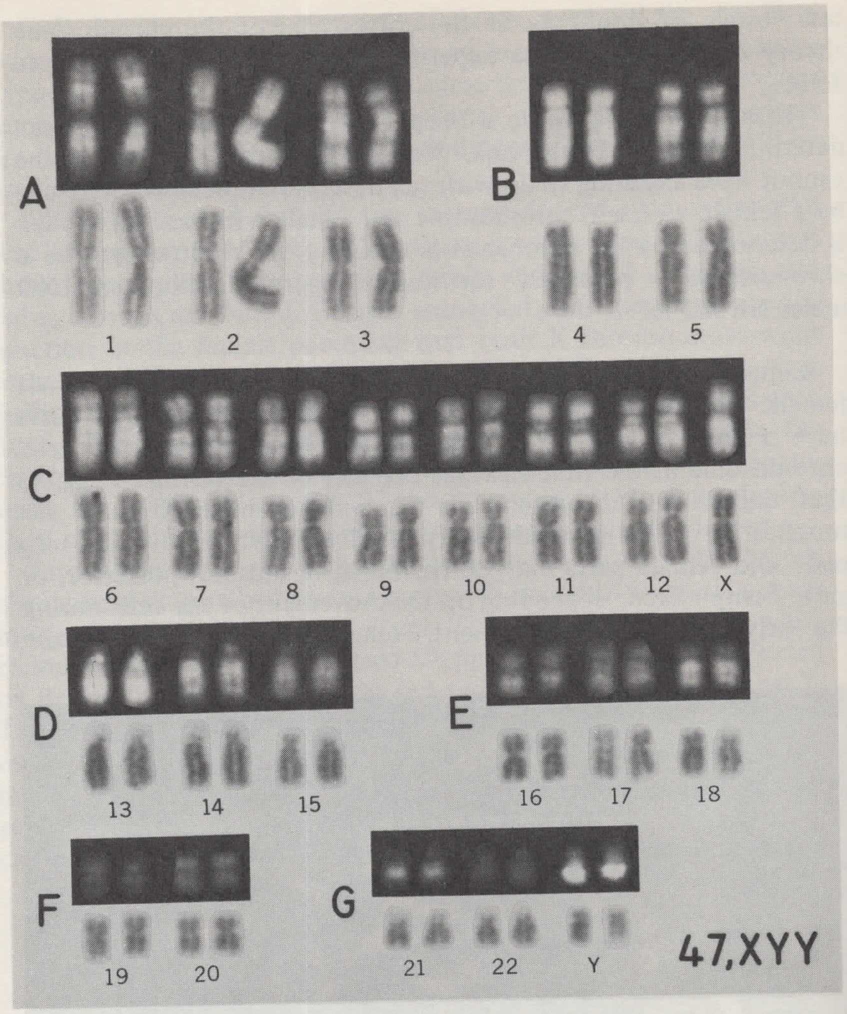


Figure 3. Human chromosomes arranged in a karyotype. Individual chromosomes from a photograph like Figure 2 are cut out and arranged by size. Human chromosomes are traditionally divided into seven size groups designated by letters A through G. The upper chromosomes are stained to show Q bands; the lower ones are stained with azure A. Each type of staining brings out characteristic landmarks that a skilled cytologist can recognize. The boy from whom these chromosomes were obtained had an extra Y chromosome.

and 1/4 male, the result of chromosome loss at the second cleavage division.

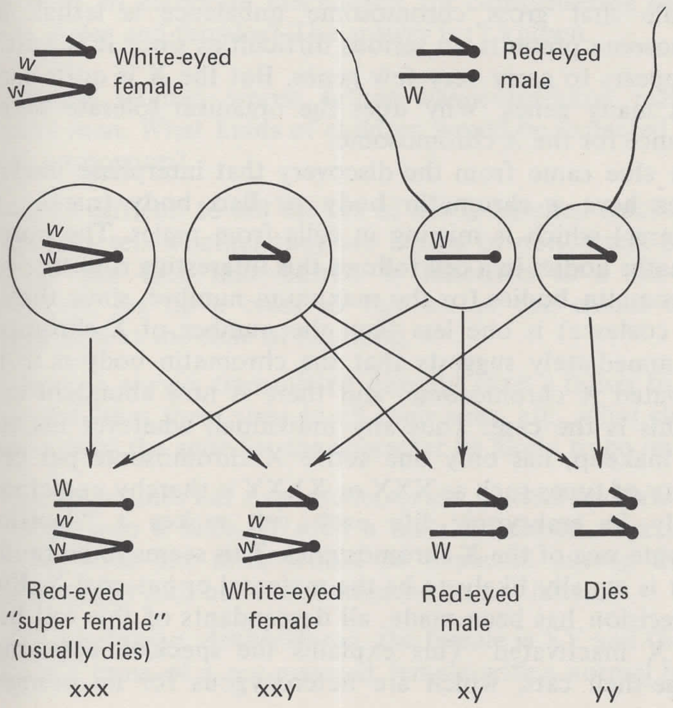
It is striking that the separation into male and female parts is very sharp. Any structure that ordinarily differs in the two sexes shows the male phenotype in the XO parts and the female type in the XX parts. For example, if the right side of the fly is male, the wing is smaller (as is characteristic of males), the abdominal stripes show the male pattern, and some bristles on the front leg (known as sex combs and found in males) are observed. The left side shows the female traits. This type of development, when each structure develops according to the genotype of its own cells and is not influenced by cells elsewhere in the body, is called **autonomous.**

Such sharp mosaicism is not found in birds and mammals. Here the development of primary and secondary sex structures is influenced by circulating chemicals, the **hormones.** A sufficiently strong hormonal influence can override the chromosome composition of a tissue. For example, in birds the ovary normally develops only on the left side, the right gonad remaining undeveloped. A number of cases are known of hens who late in life changed into males. In most of these the ovary had become diseased or had been destroyed, whereupon the previously undeveloped right gonad grew into a testis and produced enough hormone to alter the sex of the bird. Similar sex reversal can be accomplished in amphibia by grafting a testis onto a developing larva, or by hormone treatment.

ATTACHED X CHROMOSOMES

In some stocks of Drosophila one finds permanent nondisjunction of the X chromosomes, as if the two X chromosomes of the female were firmly attached to each other. This can be confirmed by examining the chromosomes with a microscope, where it is seen that the two X chromosomes are joined at the spindle fiber end. Attached-X females usually carry a Y chromosome in addition.

The diagram below illustrates the characteristic pattern of attached-X inheritance. The female is assumed to be homozygous for the recessive white-eye factor.



Of the four kinds of zygotes produced, the one with three X's and the one with no X's usually die. Thus, the two surviving types are just like their parents. This is in contrast to the usual pattern of sex-linked inheritance where the male gets his X chromosome from his mother.

SEX DETERMINATION IN OTHER ANIMALS

The Drosophila studies showing the nonimportance of the Y chromosome in sex determination, along with knowledge of the grasshopper and other animals that have no Y chromosome, led many geneticists to conclude that it is generally true throughout the animal kingdom that the Y chromosome is not the sex determiner.

However, studies in Japan on the silk moth showed the opposite result; the Y chromosome is the important factor in sex determination. XXY and XY are both females (recall that moths have XY females and XX males).

In the mouse the XO type is a normal female and the XXY is a male that is superficially normal, though sterile. Thus the Y chromosome is the sex determiner in this mammal.

In humans the XXY type is male, though sterile and abnormal in a number of respects. The condition has long been recognized medically as **Klinefelter's syndrome**. XXXY, XXYY, and XXXXY types are similar. The XO type is an underdeveloped female, known as **Turner's syndrome**. XXX humans are normal females. Thus the Y chromosome seems to be the male determining element, but the story is more complicated than in the mouse.

THE SINGLE ACTIVE X PRINCIPLE

The fact that types such as XXXYY are viable seems to violate the rule that gross chromosome unbalance is lethal. The Y chromosome presents no serious difficulties since it is quite small and appears to carry very few genes. But the X is quite large and carries many genes. Why does the organism tolerate such gross imbalance for the X chromosome?

The clue came from the discovery that interphase nuclei from females have a **chromatin body** or **Barr body** (named for its discoverer) which is missing in cells from males. The number of chromatin bodies in a cell follows this interesting rule: the number of chromatin bodies (or the maximum number, since they sometimes coalesce) is one less than the number of X chromosomes. This immediately suggests that the chromatin body is in fact an inactivated X chromosome, and there is now abundant evidence that this is the case. Thus any individual, whatever his chromosome makeup, has only one active X chromosome per cell. The viability of types such as XXX or XXXY is thereby explained.

Early in embryonic life each cell makes a "decision" to inactivate one of the X chromosomes. This seems to be random, in that it is equally likely to be the maternal or paternal X. But once this decision has been made, all descendants of this cell have the same X inactivated. This explains the speckled appearance of tortoise-shell cats, which are heterozygous for an orange color gene. The orange spots contain cells where the dark color gene is inactivated, and the dark parts have the orange gene inactivated. The size of the spot is determined by how early the decision was made. That the decision is permanent has been demonstrated with cell cultures. A cell from a woman heterozygous for genes producing two types of proteins produces only one type, and all descendants of that cell produce the same type for as long as the culture is maintained.

A woman who is heterozygous for color blindness has an average of 50% of color-blind cells in her retina, but there are enough normal cells that the effect is not noticed. The reason why females do not appear mosaic for every heterozygous X-linked gene is that most genes are not autonomous. Instead of a sharp mosaic pattern, the effect of the two kinds of cells is diffused and averaged over the body.

What is the function of this remarkable X-inactivation mechanism? It helps solve one problem: the female has twice as many X chromosomal genes as the male. On the other hand, except for those genes associated with sex, it is desirable for the genes on the X chromosome to have the same expression in males as in females. One way of doing this is to have every cell, male or female, have only one functional X chromosome. Thus there is **dosage compensation,** the same gene dosage in males as in females.

But another question is unanswered: such dosage compensation is necessary because of the inactivity of the Y chromosome. There is good evidence that originally the Y was fully functional and completely homologous to the X, but lost its activity slowly in the course of evolution. Why did this loss of function occur, necessitating such an elaborate mechanism for compensating the gene dosage in males and females? The answer isn't known.

PROBLEMS

1. A normal woman whose father was color-blind marries a normal man. What kinds of children would be expected and in what proportion?
2. It is very difficult to tell the sex of newly hatched chicks but it is easy to tell whether they are barred or nonbarred. Making use of the fact that barred is inherited as a sex-linked dominant, set up a cross so the sex of the chicks can be determined at the time of hatching.
3. If a trait is always transmitted directly from a father to all his sons and from these sons to all their sons, etc., what chromosome would the gene for the character be likely to be on?
4. A hen sometimes has a spontaneous sex reversal and late in life develops into a male. If such a male is crossed to a normal female, what sex ratio would be expected among the offspring? (Eggs with no X chromosome do not hatch.)
5. In the amphibian, *Amblystoma,* the female is XY and the male XX. In a cross of a sex-reversed female with a normal female,

Humphreys (*Am. Jour. Anat.* 76:33) found 76% females and 24% males. What does this suggest as to the viability and sex of YY animals? How could this be tested further?

6. In cats the genotype *BB* is black. *Bb* is tortoise shell, and *bb* is yellow. The gene is on the X-chromosome. A tortoise-shell female is crossed with a black male. What offspring would be expected? Would you expect to find any tortoise-shell males?

7. Occasionally a tortoise-shell male is found, and when found is sterile. Suggest an explanation for the occurrence of tortoise-shell males.

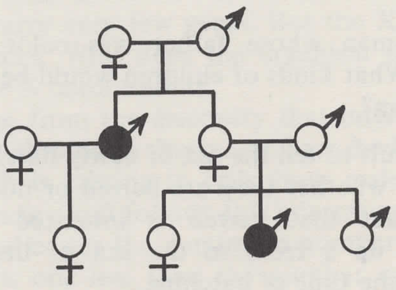
8. In humans color blindness is due to a sex-linked recessive while blue eyes are due to an autosomal recessive. Two brown-eyed persons with normal vision produce a blue-eyed color-blind son. What are the genotypes of the parents?

9. A blue-eyed woman whose father was color-blind marries a brown-eyed man whose mother was blue-eyed. What proportion of sons would be blue-eyed and color-blind?

10. In the honeybee, chartreuse eye color is due to a recessive factor, the normal brown color being dominant. A chartreuse-eyed female is mated to a normal brown-eyed male. What eye colors would be expected in the progeny?

11. What animal has a grandfather but no father?

12. Color blindness is inherited as a sex-linked recessive. Write as far as possible the genotypes of each person in the following pedigree. Black circles indicate color-blind persons.



13. A man who is color-blind has an XO daughter who has normal vision. What does this tell you about the place where nondisjunction occurred?

14. In mice there is a coat color gene called Tabby (*T*) located on the X chromosome. What genotypes would be expected among the progeny of a Tabby XO female mated with a normal gray colored (*t*) male? (At least one X chromosome is needed for the embryo to survive.)

15. Suggest
only hal
progeny

16. Explain
planting
genetic e

17. Ivory ey
offspring
for this
drone?

18. How m
XXXYY

19. If the si
single X
rather t
Suggest

20. Chromo
times. A
centrom
chances

21. Two no
Suggest

22. Turkey
out a sp
could si
the egg
chromos
the dipl
each of
female i
necessar

23. Women
fertile.
have, an

24. In the
How w

25. How mi
the chr
sperms

26. How mi
that can
the othe

27. Corn is
len are

% females and
lity and sex of
shell, and *bb* is
. tortoise-shell
pring would be
ise-shell males?
when found is
ce of tortoise-
recessive while
vo brown-eyed
ed color-blind
lind marries a
. What propor-
?
to a recessive
t. A chartreuse-
male. What eye
essive. Write as
n the following
ons.
who has normal
e where nondis-
y (*T*) located on
expected among
th a normal gray
ne is needed for

---

15. Suggest an explanation for a female Drosophila that produced only half as many sons as daughters. (Assume the number of progeny is large enough so this is not a statistical accident.)

16. Explain how it might be theoretically possible, by transplanting an ovary from one chicken to another, to have the genetic equivalent of self-fertilization.

17. Ivory eye is a recessive character in the honeybee. What offspring would be expected from a queen bee heterozygous for this gene inseminated with sperm from a normal-eyed drone?

18. How many Barr bodies (maximum) are expected in an XXXYY cell?

19. If the single active X principle is correct, an individual with a single X chromosome and no Y should be a normal female, rather than an abnormal female with Turner's syndrome. Suggest some explanations.

20. Chromosome number 21 causes mongolism when present three times. A man has two number 21 chromosomes joined at the centromere, like attached X's in Drosophila. What are his chances for a normal child?

21. Two normal parents have an XXY son who is color-blind. Suggest an explanation.

22. Turkey eggs sometimes develop parthenogenetically (i.e., without a sperm). There are three possible mechanisms: (a) the egg could simply fail to go through meiosis, remaining diploid; (b) the egg nucleus could be fertilized by a polar body; (c) the chromosomes in the egg could somehow double, thus restoring the diploid number. Predict the sex ratio in the progeny for each of these hypotheses. (Don't forget that in birds the female is XY, and assume that at least one X chromosome is necessary for the chick to live.)

23. Women with three X chromosomes are normal and fully fertile. What kinds of children would they be expected to have, and in what proportions?

24. In the kangaroo, the paternal X is inactivated in the female. How would this change the roles of X-linked inheritance?

25. How might XY and YY sperms be produced? What would be the chromosome makeup of zygotes from these kinds of sperms with normal eggs?

26. How might you explain a pair of identical twins (that is, twins that came from a single fertilized egg) in which one is male and the other is a Turner's syndrome female?

27. Corn is capable of self-fertilization since both ovules and pollen are produced by the same plant. The recessive mutant *bar-*

ren (ba) eliminates the ears, hence produces no ovules. The recessive mutant *tassel seed (ts)* converts the pollen-producing tassel into a seed producer. How could you convert corn into a bisexual species?

Robert V. Miller

es no ovules. The
e pollen-producing
convert corn into

5

PROBABILITY

The concept of probability is used to discuss events whose outcome is uncertain or statements whose truth is unknown. Probability is an expression of our degree of belief in the occurrence of an event or in the truth of a statement. For the concept to be useful in genetic analysis, we have to attach some quantitative meaning to it. If we say that the sun will probably rise tomorrow we are obviously expressing a much stronger belief than when we say that it will probably rain tomorrow. How can this degree of belief be measured?

DEFINITION OF PROBABILITY

Mathematics textbooks define probability axiomatically in terms of sets of points in a defined space. Our approach will be to think of probability as the proportion of times that an event is expected to occur in repeated trials. The proportion of occurrences in the long run is regarded as the probability of occurrence in a single trial. Our knowledge may be based on experience or it may be based on an understanding of the underlying mechanism. A situation that is particularly convenient occurs when the process can be analyzed into a number of equally likely outcomes. Then we have:

Definition: If, among n possible, equally likely outcomes there are m outcomes in which event A happens, then the probability of event A is

$$P(A) = \frac{m}{n}$$

Example 1: There are 52 possible outcomes, presumably equally likely, of drawing a card from a standard deck. Of these, four are aces. Then, by definition, the probability of an ace is 4/52, or 1/13.

Example 2: From an *Aa* x *Aa* intercross, the four outcomes *AA, Aa, aA,* and *aa* are equally likely. The probability of a dominant phenotype, *A-*, is 3/4, since there are three ways of getting this among the four possible outcomes.

Odds: The ratio of two probabilities gives the odds. In the example above, the odds of a dominant phenotype (against the alternative of a recessive) are 3/1, or 3:1, since this is the ratio of the probabilities, 3/4 and 1/4. We can always go from odds to probability by dividing the numerator of the ratio by the sum of the numerator and denominator; in this case, 3/(3 + 1) = 1/4.

THE ADDITION RULE

The addition rule states that the probability of the occurrence of either A or B (more precisely, the **union** of A and B) is the probability of A plus the probability of B minus the probability of the joint occurrence of both A and B (called the **intersection**). In symbols,

$$P(A \text{ or } B) = P(A) + P(B) - P(AB).$$

Example: The probability that a card drawn from a deck is either an ace or a spade is obviously 16/52, since there are 16 cards in the deck that are either ace or spade (or both). It is clear that the addition rule works, for

$$P = \underset{\text{Ace}}{\frac{4}{52}} + \underset{\text{Spade}}{\frac{13}{52}} - \underset{\substack{\text{Ace of}\\ \text{Spades}}}{\frac{1}{52}} = \frac{16}{52} = \frac{4}{13}$$

Most of the time in Mendelian genetics we are dealing with events that are **mutually exclusive**, meaning that the occurrence of one precludes the possibility of the other. For example, kings and aces are mutually exclusive, since a card cannot be both a king and an ace; but aces and clubs are not, because of the ace of clubs.

From the addition rule, we see that

The probability of some one of a set of mutually exclusive events is the sum of the probabilities of the single events.

Example: Returning to the Mendelian intercross, *Aa* X *Aa*, the probability of a dominant phenotype is 1/4 + 2/4, this being the sum of the probabilities of a homozygote and a heterozygote. Since the two events are mutually exclusive, their probabilities are simply additive. As a second example the probability of either a king or queen or ace is 4/52 + 4/52 + 4/52 = 3/13.

CONDITIONAL PROBABILITY

We designate by P(B/A) the probability of event B, given that event A has occurred. This is called a **conditional** probability; that is, the probability of B conditional on the occurrence of A.

Example: Suppose that a card is drawn and then, without replacing the first, a second is drawn. If the first card is a spade, the probability that the second card is a club is 13/51, since after removing a spade there remain 51 cards in the deck, of which 13 are clubs. In similar fashion, the conditional probability of another spade, given the prior occurrence of a spade and a club, is 12/50.

THE MULTIPLICATION RULE

The multiplication rule states that the probability of the occurrence of A followed by B is the probability of A plus the conditional probability of B given that A has occurred, or

$$P(A \text{ and } B) = P(A)\, P(B/A)$$

Example: The probability of an ace followed by a king in two successive drawings from a deck (without replacement) is (4/52) (4/51) = 4/663. The rule can obviously be extended to more than two events. For example, the probability that the first three cards drawn are an ace, an ace, and a king is (4/52) (3/51) (4/50) = 2/5525.

Independent Events. If the occurrence of the first event has no influence on the probability of the second, they are said to be independent. Then P(B/A) = P(B), and we have the simple rule:

> **The probability that both of two independent events will occur is the product of the individual probabilities; and likewise for more than two events.**

Example: The probability that, if a family has five children, all will be boys is (1/2) (1/2) (1/2) (1/2) (1/2) = $(1/2)^5$ = 1/32.

Here is an example that uses both the multiplication and addition rules. What is the probability that a family of five children will all be the same sex? The probability of being all boys is 1/32, and so is the probability of all girls. Since these are mutually exclusive outcomes, the probability of one or the other is the sum, or 1/16.

REPEATED TRIALS OF INDEPENDENT EVENTS

Suppose that the probability of event A is p and of event B is q, where the two events are independent and mutually exclusive, then the probability of occurrence of event A four times and event B five times in a specific order is $p^4 q^5$, by the multiplication rule. But suppose we are interested in the occurrence of this combination, four of event A and five of event B, irrespective of the order. To obtain this probability we take the probability of occurrence in a specified order and multiply by the number of orders (since any particular order has the same probability as another). The number of combinations of nine things, four belonging to one class and five to another, is

$$\frac{9!}{4!\,5!} = \frac{9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1 \times 5 \times 4 \times 3 \times 2 \times 1} = 126$$

(The symbol ! is called "factorial." A derivation of this rule is in any algebra text.) Therefore, the probability of four occurrences of event A and five of event B is

$$126\, p^4\, q^5$$

The probability rule for repeated trials of events with constant probabilities is:

> **If the probability of the occurrence of an event A is p, and the probability of event B (or the nonoccurrence of A) is q, the probability that in n trials event A will occur a total of s times and event B t times is**
>
> $$P = \frac{n!}{s!\,t!}\, p^s\, q^t$$

Notice that s + t = n and that p + q = 1. Two additional definitions are sometimes needed:

$0! = 1$ (zero factorial is one)
$X^0 = 1$ (any number raised to zero power is one)

If the binomial expression $(p + q)^n$ is expanded, the individual terms in the expansion are the same as the probability expressions given above. Therefore, this distribution of probabilities for repeated trials is often called the **binomial distribution.**

EXTENSION

The extension to more than two alternatives is direct. If the probabilities of three alternative events are p, q, and r $(p + q + r = 1)$, the probabilities of the possible outcomes are given by the terms of the expanded trinomial $(p + q + r)^n$; and similarly for more than three events.

For example, assume that the probability of a walnut-combed fowl is 9/16, rose-combed 3/16, pea-combed 3/16, and single combed 1/16—as would be the case if two doubly-heterozygous walnut-combed birds were mated. Then the probability that a dozen eggs will yield exactly six walnut-, three pea-, two rose-, and one single-combed bird (if all hatch) is

$$\frac{12!}{6!\ 3!\ 2!\ 1!}\left(\frac{9}{16}\right)^{6}\left(\frac{3}{16}\right)^{3}\left(\frac{3}{16}\right)^{2}\left(\frac{1}{16}\right)^{1} = .0254$$

PROBLEMS

1. Of the 148,462 deaths in New York in 1935, 4,196 were of diabetes and 7,436 were of tuberculosis. Based on these data, what is the probability that the next person to die will die of diabetes?

2. What is the probability that the next person to die will die of diabetes or tuberculosis?

3. Assuming a 1:1 sex ratio, what is the probability that a family of six children will consist of
   a. three boys and three girls?
   b. a boy, a girl, a boy, a girl, a boy, and a girl in this order?
   c. all girls?
   d. all the same sex?
   e. at least four girls?

4. Two parents are each heterozygous for a recessive gene for feeblemindedness. What is the probability that, if they have four children,
   a. none will be normal?
   b. all will be normal?
   c. three will be normal and one feebleminded?

5. If this family had one child which was feebleminded, what is the probability that the next child will be normal?

6. A farmer plants 100 seeds from a self-fertilized green corn

1. Two additional

wer is one)

ded, the individual
ability expressions
f probabilities for
bution.

es is direct. If the
, and r (p + q + r =
s are given by the
; and similarly for

f a walnut-combed
3/16, and single
ubly-heterozygous
probability that a
ea-, two rose-, and

0254

35, 4,196 were of
sed on these data,
to die will die of

to die will die of

ility that a family

l in this order?

ecessive gene for
hat, if they have

d?

eminded, what is
mal?

lized green corn

---

plant which is heterozygous for a recessive gene for yellow seedling color.

a. What is the probability that he will get exactly 75 green and 25 yellow plants?
b. What is the probability of his missing the expected number by less than five plants; that is, obtaining between 71 and 79 plants (inclusive)?
c. What is the probability of missing the expected number by 5 or more; that is, of getting 70 or less, or 80 or more?

Note: Don't try to compute the answers; just set the problem up. For an easy way to get an approximate answer to (c), see Appendix 1.

7. What is the probability that a family of seven children produced by the parents in problem 4 will consist of three normal girls, one normal boy, one feebleminded girl, and two feebleminded boys?

8. In Mendel's two-factor cross he obtained 315 round yellow, 101 wrinkled yellow, 108 round green, and 32 wrinkled green seeds out of 556. What is the probability of getting exactly these results? (Set up, but don't compute.)

9. *Apply a Chi-square test to Mendel's results in this experiment. Would this be considered good agreement between observation and hypothesis?

10. What is the probability that a gamete produced by a human female will contain only maternal chromosomes? What is the probability that two siblings will be genetically identical?

11. A Hereford cattle breeder has an expensive bull that he suspects may carry a recessive gene for a highly undesirable dwarfism. The breeder doesn't want to use this bull in his herd unless the bull is homozygous. He tests this by mating him with a series of cows known to have had dwarf calves and which therefore must be heterozygous. (It is impractical to use homozygous dwarf cows.) How many calves must he obtain to have "99% confidence" that the bull is homozygous—that is, so that the probability of getting this many normal calves if the bull were heterozygous is less than 0.01?

12. From the two-factor cross referred to in problem 8, what is the probability of a seed that is either yellow or wrinkled (or both)?

13. It was reported in the newspaper that one couple had had three children, all albino; and since the frequency of albinos in the population is 1/10,000, the probability of this event is $1/10^{12}$. Do you agree? If not, what is the correct probability of a sibship of three children being all albino?

*Problems with an asterisk require use of the Statistical Appendix.

14. A couple is particularly desirous of having at least one child of each sex. If they have four children, what is the probability that their wish will be granted?

15. Suppose that every couple in the world stops having children after the first son. Will this alter the sex-ratio?

16. Mendel studied seven traits. It also happens that the pea has seven pairs of chromosomes. What is the probability that no two of Mendel's traits were determined by genes on the same pair of chromosomes?

at least one child of
at is the probability

tops having children
tio?

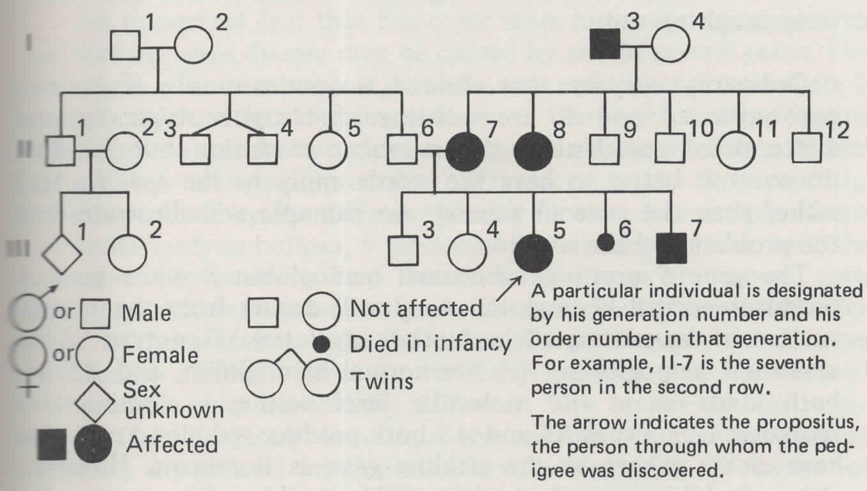
ens that the pea has
probability that no
y genes on the same

6

ANALYSIS OF HUMAN PEDIGREES

PEDIGREE ANALYSIS

Because of the long time between generations, small families, and the impossibility of making experimental matings, the study of human genetics is much more difficult than similar studies in experimental animals and plants. Nevertheless it is still possible to get considerable information from the analysis of extensive pedigrees (as well as by the application of special analytic techniques to be taken up later). A typical pedigree and a definition of the various symbols are given below.



A particular individual is designated by his generation number and his order number in that generation. For example, II-7 is the seventh person in the second row.

The arrow indicates the propositus, the person through whom the pedigree was discovered.

It is quite easy to identify some forms of inheritance from the study of pedigrees. In other cases where the inheritance is complex, the pedigrees may show no more than that the condition runs in families and offer no information as to the mode of inheritance. Some suggestions for recognizing the common types of single gene differences are given here.

1. Autosomal dominant. An autosomal dominant with regular expression has a characteristic pattern that is easy to recognize. The trait never skips a generation and a marriage between a normal and an affected individual (who is almost always heterozygous if the disease is rare) leads to a 1:1 ratio of normal to affected children.
2. Autosomal recessive. This is usually difficult or impossible to

determine on the basis of a single pedigree. In particular, it is difficult to distinguish a recessive from a dominant with a low penetrance—both result in the character skipping generations. Any of the following is suggestive of recessive inheritance: consanguinity of the parents of the affected individuals; a great majority of the affected individuals having normal parents; a 3:1 ratio of normal to affected children of heterozygous parents; all the children of two affected parents being affected; and, usually normal offspring from marriage of a normal and an affected person.

3. X-linked recessive. A much larger number of affected males than affected females is the first thing to look for. Affected males do not come from affected fathers but from carrier mothers, who may often be recognized by having affected brothers, fathers, or uncles.
4. Y-linked. Such traits if they exist at all are extremely rare. This type of inheritance is characterized by having only males showing the trait. Furthermore, all the sons of an affected person are affected.
5. X-linked dominant. This is also very rare in man. Pedigrees show affected males with normal wives transmitting the trait to daughters only.

THE USE OF THE WORDS DOMINANT AND RECESSIVE

Ordinarily, we say that allele *A* is dominant to *a* if the two genotypes *AA* and *Aa* are indistinguishable phenotypically, and different from *aa.* Usually this causes no confusion, but if it does do so it is better to have the words apply to the specific trait rather than the gene in general. An example will illustrate both the problem and the solution.

The gene *S* produces abnormal hemoglobin S while gene *A* produces normal hemoglobin A. (I will depart from the normal custom of designating alleles by the same letter.) Genotype *SS* has abnormal hemoglobin, *AA* has normal hemoglobin, and *AS* has both kinds—so at the molecular level neither is dominant or recessive. Genotypes *SS* and *AS* both produce red blood cells that have sickle shapes so the sickling *gene* is dominant. However, although *AS* is normally healthy, *SS* has sickle cell anemia, so the *gene* is recessive. Thus, whether the gene *S* is dominant, recessive, or neither depends on whether the trait under discussion is the presence of sickle cells, anemia, or the kind of hemoglobin present. If we speak of the dominance of the gene with respect to the specific trait, we avoid any semantic confusion. When there is danger of ambiguity I will follow this usage.

A problem that frequently arises in human genetics is that we often do not know the phenotypes of all three genotypes. For example, if the genotype *Aa* causes a rare disease and *aa* is normal it is likely that the phenotype of *AA* is not known because the *Aa* x *Aa* mating has not occurred. It is the custom in human genetics to speak of a trait as dominant if it is expressed in the heterozygote whether the homozygous phenotype is the same or not.

GENETIC PROGNOSIS

One use that can be made of genetic knowledge is advising prospective parents about the chances of hereditary disease in their children. For example, parents who have produced a child with a disease known to be caused by a recessive factor can be advised that the chances are one in four for any future child to have the disease. Whether the parents would choose to have another child under these conditions would depend on the severity of the disease and their willingness to take the chance.

Under most circumstances a genetic advisor can do no more than state probabilities. Sometimes more definite statements can be made. For example, it is possible to tell a woman who comes from a family that is free of hemophilia (due to a sex-linked recessive) but whose husband has a hemophilic brother that there is virtually no possibility of her producing a child with hemophilia. Likewise, any normal person in a family in which there is a disease caused by a dominant gene will not ordinarily transmit the disease.

Another use that can be made of genetic information is early diagnosis of disease. Knowledge of a family history of some disease enables the physician to be on the alert for the slightest sign of the disease and thus institute preventive or curative procedures at the most advantageous time.

An important fact that has come from human genetics research is that the same disease may be caused by any of several genes. For example, albinism is usually due to a recessive factor (and there is some evidence that there may be more than one recessive factor) but may be caused by a dominant gene. In some diseases a careful study has revealed detectable differences in the forms of the disease caused by the different genetic factors. An example of this is epidermolysis bullosa, a tendency for blisters to form with the slightest provocation. Two forms of the disease are known, each due to a different dominant gene. They can be differentiated since one is more severe than the other. Often, though, the diseases caused by different genes are clinically identical, and the mode of inheritance may be entirely different. This means that in many cases it is necessary to obtain the individual's pedigree before making any genetic prediction rather than inferring the manner of inheritance from a study of similar diseases in other families.

GENETIC COUNSELING

In recent years there has been increased emphasis on genetic abnormality and disease. Part of the reason is the control over infectious agents, which used to cause most diseases. Another reason is the much greater knowledge of inherited disease. The most recent edition of V. A. McKusick's list of known Mendelian traits in man lists 583 known dominants (along with 635 more that are less well established), 466 recessive (and 481 less certain), and 93 X-linked recessives (plus 78 less certain). They range from normal and trivial defects to the most severe diseases and affect every part of the body. Most are rare, but collectively they add up to a great deal of human suffering.

A genetic counselor is usually a physician who specializes in inherited diseases. From a careful diagnosis, knowledge of the inheritance of this condition, and examination of the pedigree he is often able to give the probability of an affected child from a specified marriage. If the trait follows simple Mendelian inheritance this is usually not difficult to calculate.

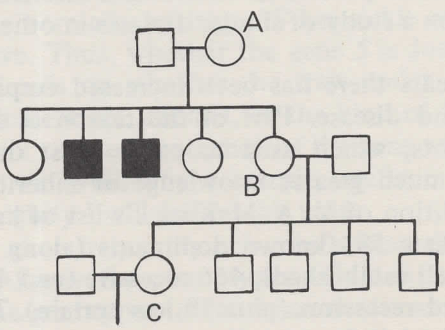
For more complicated inheritance, it may be necessary to rely on empirical data. The probabilities of being affected, based not on genetic theory but on past experience, are called **empiric risk** figures. Although not as satisfying as probabilities based on a thoroughly understood theory, they are of great practical utility.

For more and more diseases which are inherited as recessives, ways are being found to detect heterozygous carriers of the causative gene. Sometimes the homozygote can be detected as an embryo. Thus, for **Tay-Sachs disease**, which causes the child to degenerate and die in the first few years of life, the carriers can be detected by a simple blood enzyme test. Then if both parents are heterozygotes, and the risk of an affected child is 1/4, the amniotic fluid from each embryo can be examined and the abnormal embryos aborted. For other diseases where this is not possible, artificial insemination from a noncarrier male is sometimes used.

Genetic counseling is now becoming a regular part of most large medical centers.

USE OF CONDITIONAL PROBABILITIES IN GENETIC COUNSELING: BAYES' FORMULA

From knowledge of the inheritance of a disease and information about a person's ancestry we can often make probability statements about his genotype and use this for genetic prognosis. For example, in the accompanying pedigree of **Lesch-Nyhan disease**, one of the most tragic X-linked conditions, we are asked by woman C what the chances are that if she has a son he will inherit the disease that her uncles had. We know from the fact that A produced two diseased sons that she is heterozygous and carries the abnormal gene. We also know that there is a probability 1/2 that her daughter B inherited the gene; furthermore, there is a



probability 1/4 of C inheriting the gene. But can we use the information that C had four normal brothers to tell us more about her genotype? Having four normal sons is much more likely if B is homozygous than if she is heterozygous. We can use the information in this way: let N and n stand for normal and abnormal genes, respectively.

Genotype of B	*Nn*	*NN*
Prior probability of B's genotype	1/2	1/2
Conditional probability of four normal sons, given B's genotype	1/16	1
Product of the two above	1/32	1/2
Ratio of products = posterior odds for B's genotype	1 :	16
Posterior probability of B's genotype	1/17	16/17

Since she produced four normal sons, the odds are 16:1 that B is NN rather than Nn. The chance of her being a carrier has been reduced from 1/2 to 1/17 by the additional information provided by the four normal sons. Therefore, the probability that C is a carrier is 1/34, and if she has a son, his chances of having the Lesch-Nyhan disease is 1/68, or less than 1.5%.

This is an example of Bayes' formula. The ratio of the two conditional probabilities is usually called the *likelihood ratio.* So we can state Bayes' rule in these words:

Posterior odds = prior odds X likelihood ratio

In the example just discussed, the posterior odds that B is heterozygous are

$$\text{Posterior odds} = \frac{1}{1} \times \frac{1}{16} = 1{:}16$$

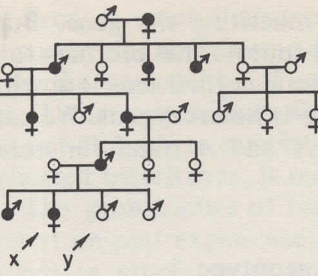
It may bother you that the probability of B being heterozygous changed as she produced additional normal sons. In one sense this is ridiculous for we know that whatever her genotype actually was, it didn't change when she produced the sons. If we had complete knowledge, we would know which genotype she actually is, in which case the probability of this would be one and of the alternative, zero. But we are talking about probabilities based on whatever information we can bring to bear on the subject, and for this the fact that she produced four normal sons changed our information.

PROBLEMS

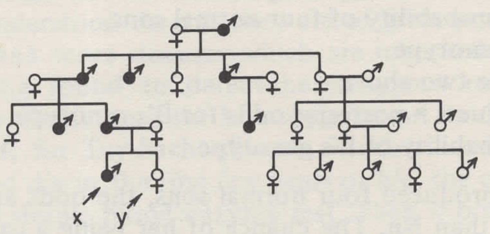
In the following pedigrees assume that the abnormal trait is rare in the population. Answer the following questions about each pedigree:

- What is the probable mode of inheritance?
- What is the probability of the first child being affected if individual x marries a normal unrelated person?
- What is the probability of the first child being affected if individual y marries a normal unrelated person?

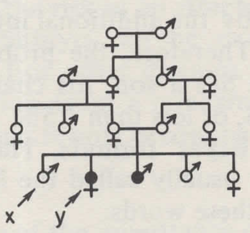
1.



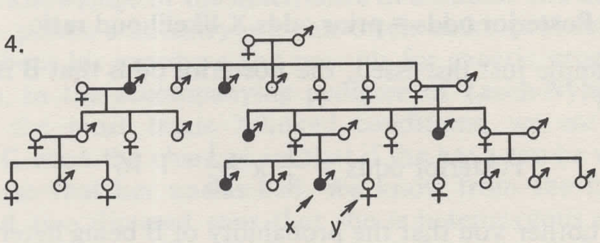
2.



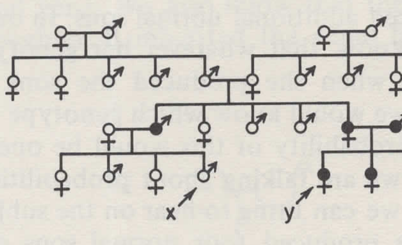
3.



4.



5.



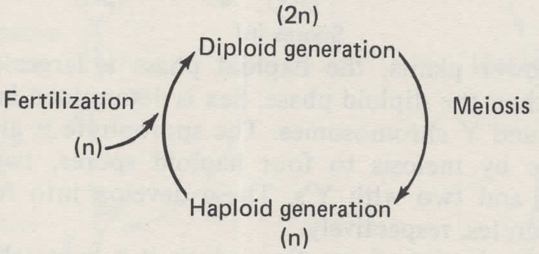
6. In pedigree 5 what is the probability of an abnormal child if individual x marries an affected woman?
7. In pedigree 3 what is the probability of an abnormal child if individual x marries an affected woman?
8. About 1 normal person in 70 is heterozygous for albinism. A normal woman whose parents were normal but who has an albino brother marries an unrelated normal man. What is the probability that if they have a child it will be an albino? Would the probability be larger if the woman marries her cousin? (Assume that albinism is due to a recessive gene.)

7

LIFE CYCLES

THE GENERAL PATTERN OF LIFE CYCLES

There is a general pattern of life cycles that is found throughout the animal and plant kingdoms. There is a diploid generation, consisting of a series of mitotic divisions of diploid cells, followed by meiosis. Then comes a series of haploid divisions, followed by fusion of two haploid nuclei (fertilization) to give the diploid again. This can be diagrammed as follows:



In different species there may be abbreviations, or even omissions of one part of the cycle. For example the algae and fungi generally have a reduced diploid phase while the higher plants have a shortened haploid phase. In plants it is conventional to call the haploid phase the **gametophyte** and the diploid the **sporophyte**.

Some animals and plants reproduce entirely asexually, so that the progeny are a genetic carbon copy of the parent.

THE SIGNIFICANCE OF SEXUAL REPRODUCTION

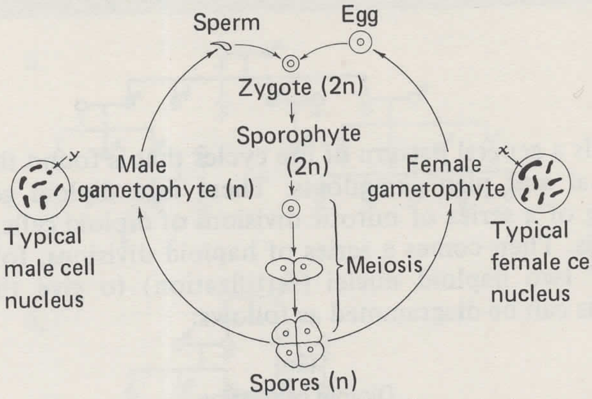
What can be the evolutionary advantage of biparental reproduction? It is not particularly efficient as a purely reproductive mechanism; for sheer numbers of progeny some asexual species are the best.

The key to the significance is provided by Mendelian inheritance. The process of meiosis and fertilization permits Mendelian assortment of the genes in every generation, and in this shuffle various new combinations can be tried out by the species. In an asexual organism if two favorable mutations occur in different individuals, the two mutant forms can only compete with each other. In a sexual form they can be recombined so that one individual can have both. This permits an incessant trial-and-error mechanism by which different mutant combinations occur and the more favorable ones are retained by the sieve of natural selection while the ill-adapted ones are eliminated.

Asexual forms may have an immediate advantage because they don't produce any poor new combinations. On the other hand, they are at a disadvantage in the long run, when environmental conditions change.

In the following sections, life cycles of various types will be illustrated with several organisms that are of particular interest in genetics research.

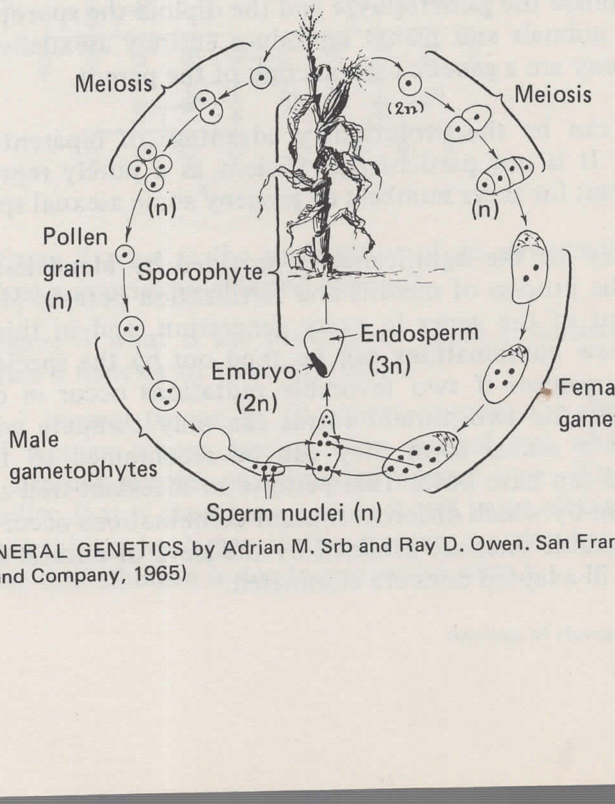
SPHAEROCARPOS, A LIVERWORT


Sperm
Egg
Zygote (2n)
Sporophyte
(2n)
Male
gametophyte (n)
Female
gametophyte
Y
X
Typical
male cell
nucleus
Typical
female cell
nucleus
Meiosis
Spores (n)

In many lower plants, the haploid phase is larger and more conspicuous than the diploid phase. Sex is determined in Sphaerocarpos by X and Y chromosomes. The sporophyte is always XY, and gives rise by meiosis to four haploid spores, two with X chromosomes and two with Y's. These develop into female and male gametophytes, respectively.

From the standpoint of genetic analysis it is especially convenient that the four spores produced by meiosis are held together. The 1:1 segregation ratio of the two gamete types from a heterozygote in most organisms can only be determined statistically.

MAIZE, AN EXAMPLE OF A HIGHER PLANT


Meiosis
(2n)
Meiosis
(n)
(n)
Pollen
grain
(n)
Sporophyte
Endosperm
(3n)
Embryo
(2n)
Female
gametophytes
Male
gametophytes
Sperm nuclei (n)

(after GENERAL GENETICS by Adrian M. Srb and Ray D. Owen. San Francisco: W. H. Freeman and Company, 1965)

ecause they
other hand,
vironmental

pes will be
r interest in

cell

r and more
in Sphaero-
always XY,
wo with X
female and

ally conven-
ld together.
pes from a
ned statisti-

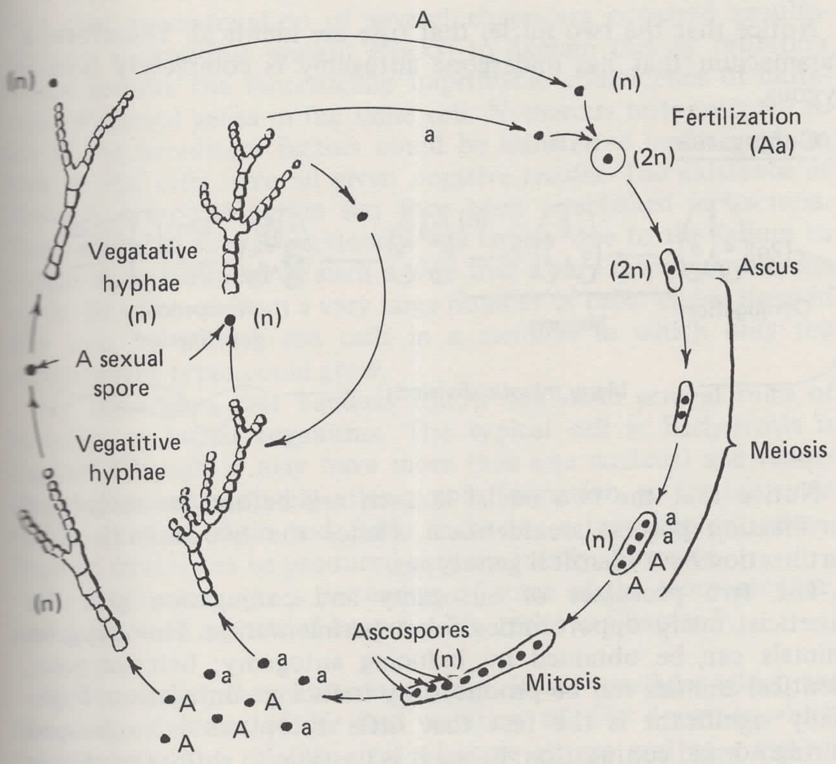
is

male
metophytes

rancisco: W. H.

Here a sporophyte phase is expanded and the gametophyte is reduced to three nuclei in the male and eight in the female. In addition to the usual fertilization to produce the diploid embryo, there is also a peculiar fusion of two nuclei from the female gametophyte with one from the male, to produce a 3n (triploid) that develops into the **endosperm**. The endosperm is the nutritive part of the seed and is no longer functional after the seed germinates. Since the two nuclei from the female gametophyte are derived by mitosis from a single nucleus, the endosperm always contains two identical genes from the female and one from the male parent.

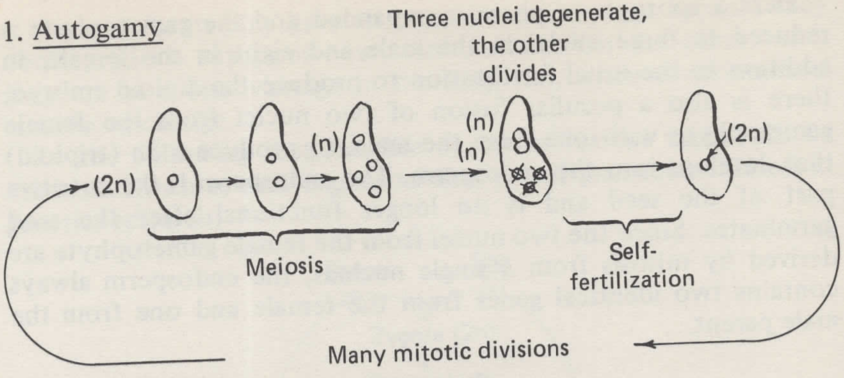
NEUROSPORA, A FUNGUS



In Neurospora, the gametophyte (n) is the dominant phase. Meiosis occurs immediately after fertilization, so the sporophyte (2n) phase is virtually eliminated. The haploid gametophytes (known as hyphae) are of two types, *A* and *a*. Matings occur only between opposite types; therefore the zygote is always heterozygous. Meiosis follows immediately giving rise to four spores, two carrying *A* and two carrying *a*. For genetic analysis it is especially convenient that the spores produced by meiosis not only are held together in the **ascus**, but are in order. Spores number 1 and 2 separated from 3 and 4 at the first meiotic division, whereas 1 and 2 separated from each other, and 3 and 4 from each other, at the second meiotic division. Each of the spores divides once again by ordinary mitosis, so there are eight haploid spores. Each group of two consists of two identical spores.

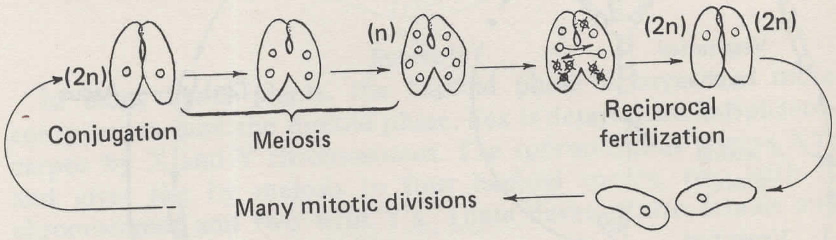
PARAMECIUM, ONE OF THE PROTOZOA

1. Autogamy


Three nuclei degenerate, the other divides
(2n)
(n)
(n)
(n)
(2n)
Meiosis
Self-fertilization
Many mitotic divisions

Notice that the two nuclei that fuse are identical. Therefore, a Paramecium that has undergone autogamy is completely homozygous.

2. Conjugation


(2n)
(n)
(2n)
(2n)
Conjugation
Meiosis
Reciprocal fertilization
Many mitotic divisions

Notice that the two nuclei in each cell before the reciprocal fertilization process are identical. Hence the two animals after fertilization have identical genotypes.

The two processes of autogamy and conjugation give the geneticist many opportunities for experimentation. Homozygous animals can be obtained by inducing autogamy; heterozygous, identical animals can be produced by inducing conjugation. Especially significant is the fact that little cytoplasm is exchanged during normal conjugation; hence it is possible to study the effects of identical nuclei in different cytoplasms.

The nucleus that has been discussed here is the micronucleus. Paramecium also has a larger macronucleus, but it has not been considered since it does not enter into the sex cycle.

ESCHERICHIA COLI, A BACTERIUM

Until 1946 there was no evidence of a sexual phase in the bacterial life cycle, but Mendelian recombination has now been found in a strain of *Escherichia coli.* A wild type of *E. coli* will grow on a minimal medium containing no vitamins or amino acids. By treatment with X rays and ultraviolet rays, mutants were obtained which would not grow unless certain amino acids or vitamins were added to the medium. E. L. Tatum and J. Lederberg in 1946 were able to show recombination in these strains by means of an experiment, the essentials of which are given here.

The two strains used were:

Strain I - *l M* (requires the amino acid leucine for growth)
Strain II - *L m* (requires the amino acid methionine for growth)

The two strains were mixed and placed on a minimal medium containing neither leucine nor methionine. About one cell in a million gave rise to a colony. These cells must have been of genotype *L M* since the medium permitted growth of this type only.

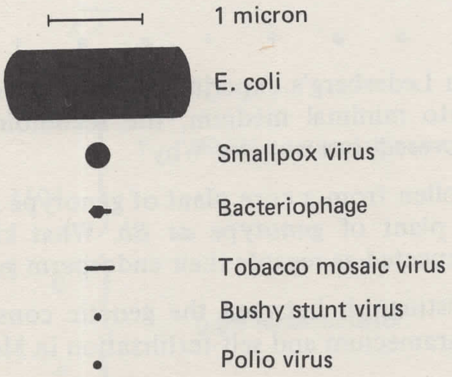
It is difficult to account for these nutritionally independent *L M* cells by any process other than that of sexual fusion of *l M* and *L m* cells. Mutation was ruled out as a possible explanation by the fact that recombination of several characters occurred simultaneously in multiple mutant stocks; to explain this by mutation would require the fantastically improbable coincidence of mutation of several genes in the same cell. Numerous tests designed to see if the hereditary factors could be transferred in the medium free of the cells have all given negative results. The existence of Mendelian recombination has thus been established in bacteria. The failure to find it previously was largely due to the failure to design an experiment in such a way that a very few recombinants could be selected from a very large number of cells, as was done in this case by placing the cells in a medium in which only the recombinant types could grow.

We have seen that bacteria follow the same general rules of heredity as higher organisms. The typical cell in Escherichia is haploid (though it may have more than one nucleus) and reduction follows immediately after zygote formation, so the life cycle is like that of algae and fungi in having a long haploid phase. Diploid strains can be produced experimentally, however.

We shall return to a discussion of some of the more peculiar, and interesting aspects of bacterial life cycles in Chapter 9.

BACTERIOPHAGE, A VIRUS

Much smaller than bacteria are the viruses—particles which are the simplest forms of life that we know of. The drawings which follow show the relative sizes of some viruses and of a typical bacterial cell for comparison.



The virus that has had the most genetic study is one of the bacteriophages. It is tadpole shaped and infects the bacterial cell by attaching tail first and injecting its contents into the bacterial cell. A half hour or so later the bacterial cell erupts (lyses) and dozens of bacteriophages come out, each of which goes on to infect another bacterial cell. This cycle continues until all the available bacterial cells have been destroyed.

Two convenient types of mutations in phages are: (1) those affecting the range of bacterial hosts that can be infected, and (2) plaque-type mutations. When a phage particle is placed on a "lawn" of bacterial growth an area of lysed cells is formed in a few hours. The size and appearance of this "plaque" is characteristic of the virus strain, and mutations changing the plaque type are easily discerned. The phenomenon of mutation in phages obeys the same rules as mutation in higher organisms. Mutations also occur in other types of viruses, such as those causing various plant and animal diseases.

Recombination may be shown by infecting bacterial cells with two types of phage. For example, *h* and *H* cause different host specificities and *r* and *R* different sized plaques. When strain 1 (*h r*) and strain 2 (*H R*) phages infect the same cell four types of particles are released:

H R same as strain 2
h r same as strain 1
$\left.\begin{matrix} Hr \\ hR \end{matrix}\right\}$ recombinant types

PROBLEMS

1. A Paramecium of genotype $A/a/B/b$, where *A* and *B* are independent factors, undergoes autogamy. What are the possible genotypes, and what is the probability of each?
2. One of the following ascospore arrangements does not occur normally in Neurospora. Which is it?

a. *a a* + + + + *a a*
b. *a a a a* + + + +
c. *a* + + *a* + + *a a*
d. *a a* + + *a a* + +

3. In Lederberg's experiment, when *l M* and *L m* bacteria were put into minimal medium, the recombination type *L M* was recovered, but not *l m*. Why?
4. Pollen from a corn plant of genotype *Aa bb* is put on the silk of a plant of genotype *aa Bb*. What kinds of kernels would be expected as regards their endosperm genotypes?
5. Distinguish between the genetic consequences of autogamy in Paramecium and self-fertilization in Mendel's peas.

one of the
acterial cell
he bacterial
(lyses) and
goes on to
ntil all the
: (1) those
ted, and (2)
laced on a
ed in a few
acteristic of
e are easily
ys the same
cur in other
and animal
l cells with
fferent host
strain 1 (*h*
ur types of
d *B* are in-
he possible
not occur
a were put
M was re-
the silk of
would be
togamy in

LINKAGE AND CHROMOSOME MAPPING

8

LINKED GENES IN CHICKENS

In all the crosses that have been considered so far we have assumed that different pairs of factors are inherited independently of each other. This is true when the genes are on different chromosomes, but when two genes are on the same chromosome they tend to stay together in inheritance, a phenomenon known as **linkage**. This can be illustrated by some data from chickens.

In these crosses the following genes were involved:

C - creeper (short legs, lethal when homozygous)
c - normal
R - rose comb
r - single comb

A homozygous rose-combed, normal strain was crossed with a single-combed, creeper strain. The creeper hybrids were test-crossed with a single-combed, normal strain. The total offspring produced from many such crosses were: 1069 rose-combed, normal; 1104 single-combed, creeper; 6 rose-combed, creeper; and 4 single-combed, normal. If these factors were following the rule of independent assortment the four classes of offspring would be expected in equal numbers. But in this case most of the offspring have the same combinations of genes as the original parents from which the first hybrids were produced.

The cross may be diagrammed as follows:

$$\frac{R\ c}{r\ C} \times \frac{r\ c}{r\ c}$$

$$\left.\begin{array}{ll} \dfrac{R\ c}{r\ c} & 1069 \\ \dfrac{r\ C}{r\ c} & 1104 \end{array}\right\} \text{99.5\% nonrecombinants}$$

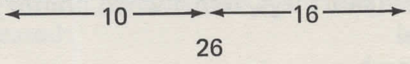
$$\left.\begin{array}{ll} \dfrac{R\ C}{r\ c} & 6 \\ \dfrac{r\ c}{r\ c} & 4 \end{array}\right\} \text{0.5\% recombinants}$$

The last two kinds of offspring could be produced only by some kind of exchange of parts between the chromosomes in the heterozygous parent. Such a process is known to occur and is called **crossing-over.**

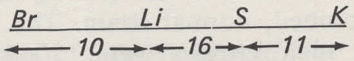
CHROMOSOME MAPPING

If crossing-over occurs at various points along the chromosome more or less at random, two genes should be separated by crossing-over very rarely if they are close together but more frequently if they are far apart. If crossing-over between genes *a* and *b* is twice as frequent as between *c* and *d,* then it is a reasonable hypothesis that *a* and *b* are twice as far apart on the chromosome as are *c* and *d.* Furthermore, it should be possible to arrange the various genes in order according to their crossover distances.

In chickens there is 10% recombination between the genes for brown eye (*Br*) and light down (*Li*). There is 26% between brown eye and silver plumage (*S*), and 16% between silver plumage and light down. There is only one arrangement of the genes compatible with these figures, that is



Another gene, slow feathering (*K*) is found to have 11% recombination with S and 27% with *Li.* The arrangement of all four genes therefore is



In this fashion a gene map of an entire chromosome may be made up. The map distances are measured in units of 1% crossing-over, so if two genes are 15 units apart there is an average of 0.15 crossovers between them. The distances represented on a chromosome map are proportional to the actual physical distances on the chromosome only if crossing-over occurs with equal frequency in all parts of the chromosome, which in fact is not the case. But the order of the genes is correct and the distances approximately so.

The idea that genes are arranged in a row along the chromosome and that their distance apart could be measured in terms of crossover values was conceived and developed by the Morgan school of Drosophila workers between 1910 and 1916. The key idea came from A. H. Sturtevant. Since then the idea has been proved to be correct by many experiments on a variety of experimental animals and plants.

A CROSS INVOLVING THREE PAIRS OF FACTORS

In a series of Drosophila crosses involving three pairs of sex-linked genes the results on the next page were obtained (data from Bridges and Olbrycht. 1926. *Genetics* 11:41). The mutant genes involved are: *sc* - scute (scutellar bristles missing), *ec* -

echinus (rough eye), *cv* - crossveinless (wing crossveins missing). As has become standard practice, the wild-type alleles of each of the mutants will be designated simply as "+".

From the cross

$$\frac{+\quad\overset{\text{region 1}}{}\quad +\quad\overset{\text{region 2}}{}\quad +}{sc\qquad ec\qquad cv}\ ♀\ \times\ \underline{sc\qquad ec\qquad cv}\ ♂$$

The offspring obtained were:

GENOTYPE Male	Female	PHENOTYPE	NUMBER		
$\underline{sc\ ec\ cv}$	$\frac{sc\ ec\ cv}{sc\ ec\ cv}$	Scute, echinus, crossveinless	1158		From non-crossover female gametes
$\underline{+\ +\ +}$	$\frac{+\ +\ +}{sc\ ec\ cv}$	Wild type	1455		
$\underline{sc\ +\ +}$	$\frac{sc\ +\ +}{sc\ ec\ cv}$	Scute	163	9.0%	From single crossovers region 1
$\underline{+\ ec\ cv}$	$\frac{+\ ec\ cv}{sc\ ec\ cv}$	Echinus, crossveinless	130		
$\underline{sc\ ec\ +}$	$\frac{sc\ ec\ +}{sc\ ec\ cv}$	Scute, echinus	192	10.5%	From single crossovers region 2
$\underline{+\ +\ cv}$	$\frac{+\ +\ cv}{sc\ ec\ cv}$	Crossveinless	148		
$\underline{sc\ +\ cv}$	$\frac{sc\ +\ cv}{sc\ ec\ cv}$	Scute, crossveinless	1	0.1%	From double crossovers regions 1 and 2
$\underline{+\ ec\ +}$	$\frac{+\ ec\ +}{sc\ ec\ cv}$	Echinus	1		
			3248		

The flies with the scute phenotype, for example, must have come from a female gamete of the constitution *sc* + +, which requires an exchange in region 1. Similarly the echinus and the scute, crossveinless types require double crossovers, and so on.

From these data we infer that the distance from *sc* to *ec* (region 1) is 9.0 + 0.1 or 9.1 units, since this is the total percent of crossovers in this region. Likewise the region from *ec* to *cv* (region 2) is 10.5 + 0.1 or 10.6 units. The map for these genes is

sc —— 9.1 —— *ec* —— 10.6 —— *cv*

INTERFERENCE

An advantage of a cross involving three factors at once is that it gives just as much information in a single experiment as would be obtained in three experiments in which two factors were considered at a time. It is possible to carry out experiments in which as many as six or eight gene loci are studied at once, and this is frequently done in Drosophila. This makes possible the detection of triple and quadruple crossovers. By accumulation of such data the standard chromosome maps such as are published have been worked out.

In the example above there was a crossover frequency of .091 in region 1 and .106 in region 2. If these represent the probabilities of crossovers in the two regions, and if these are independent, then the probability of a double crossover is (.091) (.106) or .010. The proportion of doubles actually observed was only .001 or one tenth the expected amount. Almost always in experimental data the proportion of double crossovers is less than would be expected on a random basis. The occurrence of one crossover apparently interferes with the occurrence of other crossovers in adjacent regions, and this phenomenon is known as **interference.**

The observed number of double crossovers divided by the expected number is the **coincidence.** There are several ways in which interference is measured, but the simplest is

$$\text{Interference} = 1 - \text{coincidence}$$

DETERMINATION OF GENE ORDER IN A THREE FACTOR CROSS

If the order of the genes on the chromosome is not known, this can readily be determined. Consider the testcross between the triple heterozygote (*RMC*) / (*rmc*) and the recessive homozygote (*rmc*) / (*rmc*), where the order of the genes within the parentheses is not known. Twelve-hundred and eighty progeny were counted and classified, leading to the results below. Only the chromosome from the heterozygous parent is given, since the contribution from the homozygous parent is always *rmc*.

RMC	413	*Rmc*	170	
rmc	426	*rMC*	161	
RMc	6	*RmC*	47	Total = 1280
rmC	3	*rMc*	54	

One way to determine the gene order is from the double crossover class. From the small numbers, six and three, we know that *RMc* and *rmC* are double crossover chromosomes. Notice that if *C* and *c* are interchanged the noncrossover gametes are obtained. Thus *C* must be in the center and the gene order is *RCM* or *MCR*. Once the gene order is known, the procedure of the previous page can be followed.

Another method is to regard the gene pairs two at a time. Ignoring *C* and *c*, the number of gametes involving recombination

Between *R* and *M* is 170 + 161 + 47 + 54 = 432, or 33.8%

Likewise, between *R* and *C* is 6 + 3 + 170 + 161 = 340, or 26.6%

Finally, between *C* and *M* is 6 + 3 + 47 + 54 = 110, or 8.6%

Clearly *R* and *M* are the most distant and the map is

R *C* *M*
26.6 8.6

Notice that 110 + 340 does not equal 432. This is because of double crossovers, of which there are 9, or 0.7%.

The coincidence is (.007) / (.266) (.086) = .3; the interference is .7.

ABSENCE OF CROSSING-OVER IN THE DROSOPHILA MALE

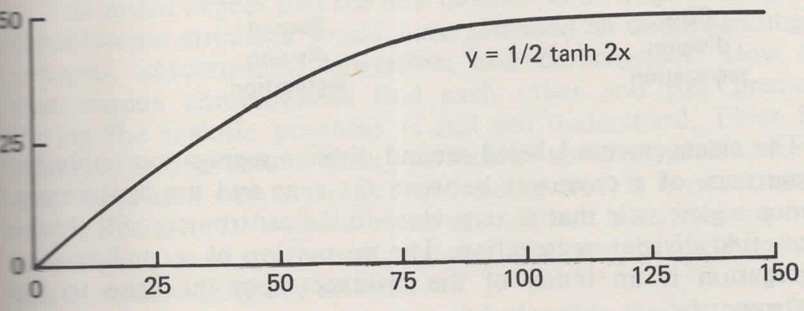
In most animals and plants that have been studied genetically, crossing-over occurs in both sexes. This is also true of the human species, for although not very many unequivocal instances of crossovers are known, they include crossovers in both sexes.

On the other hand, Drosophila is a conspicuous exception to this. For some reason that is not at all understood, there is no crossing-over in the male Drosophila. A similar instance is found in the silk moth. Here, it is the female in which crossing-over is absent, but recall that the female is the XY sex in moths.

RELATION BETWEEN RECOMBINATION PERCENT AND MAP DISTANCE

If a crossover test is made involving two pairs of genes that are far apart on the chromosome, there will be some undetected multiple crossovers. This means that the recombination percent between the two distant genes will be an underestimate of the map distance. Two genes that are so far apart on the chromosome as to be independent of each other will have a recombination value of 50%. The proper correction to make would depend on the amount of interference, and this is not constant. However, a convenient and frequently used formula is one by Kosambi, and shown in the drawing below. The ordinate (100 y) is the recombination percent; the abscissa (100 x) is map distance.

MAPPING FUNCTION
(after Kosambi)



The equation, 2y = tanh 2x (where tanh stands for the hyperbolic tangent), was derived from the assumption that the coincidence is proportional to the recombination fraction, y.

THE TIME AND PLACE OF CROSSING-OVER

During the prophase of meiosis the homologous chromosomes synapse (i.e., come to lie side by side). At such a time the breakage and reattachment of the crossover process could occur. There is evidence from a number of sources that a crossover involves only two of the four strands of chromosomes.

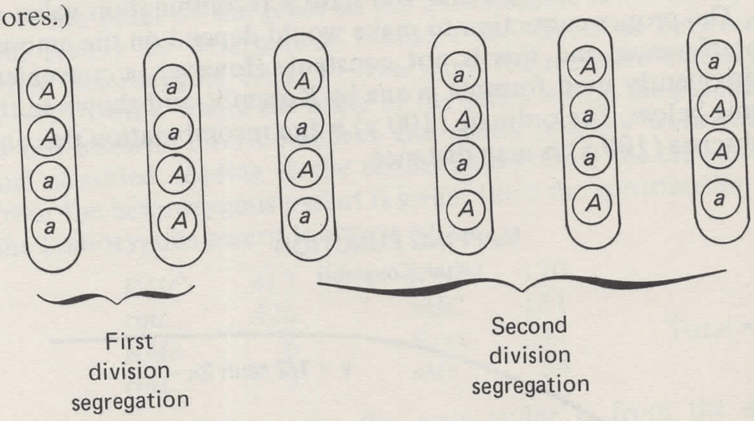
The most direct evidence comes from some of the lower plants, such as Neurospora, where the four cells that are derived from a single meiosis are held together. When these are observed, it is invariably noted that whenever there is a crossover only two of the four spores carry recombinant chromosomes, the other two carrying the original chromosomes. This means that only two of the four chromatids were involved in the crossover.

Similar evidence is derived from nondisjunctional eggs in Drosophila. One occasionally finds that of two chromosomes that come from a single female, one will be a crossover and the other a noncrossover, which would show that the crossover event had involved only one of these two chromosomes.

When a double crossover occurs, the second crossover may involve the same two strands as the first, or the other two, or one from each group. These are called, respectively, two-strand, four-strand, and three-strand double crossovers.

MAPPING GENES WITH RESPECT TO THE CENTROMERE IN NEUROSPORA

In Neurospora the ascospores occur in definite orders in the ascus. The following orders are possible from a plant heterozygous for a pair of genes, *A* and *a*. (Only four spores are shown, since each group of two adjacent spores is made up of two identical spores.)



The arrangements labeled second division segregation imply the occurrence of a crossover between the gene and the centromere. Hence a gene pair that is very close to the centromere will always show first division segregation. The proportion of second division segregation is an index of the distance from the gene to the centromere.

Since only half of the chromosome strands are actually involved in any crossover, the recombination frequency is only one half the proportion of second division segregations.

Thus, for genes so close to the centromere that double crossovers (and, of course, multiple crossovers higher than double) can be ignored, the map distance from the centromere, d, is given by

$$d = \frac{1}{2} \text{ percentage of second division segregation asci.}$$

THE SIGNIFICANCE OF CROSSING-OVER

If there were no crossing-over, two linked genes would be inseparably attached. If, for example, a poultry breeder had a creeper gene linked with a gene for rose comb he would not be able to get any such thing as a creeper, single combed or a noncreeper, rose combed bird. In the evolutionary history of a species crossing-over is important because it permits genes on the same chromosome to participate in the same shuffling and reassortment that genes on different chromosomes do, thus extending the evolutionary benefits of sexual reproduction to genes on the same chromosome.

A RECONSIDERATION OF MEIOSIS

The account of meiosis given in Chapter 3 was quite superficial. Now that crossing-over, which is an important aspect of meiosis, has been discussed, it is worthwhile to consider the process again in more detail. The diagrams at the end of this chapter show the various meiotic stages. The diagrams on the left show the entire cell very roughly as it would look in a microscope; those on the right show the chromosomes in more detail. The black and white colors designate chromosomes of paternal and maternal origin.

RECOMBINATION WITHIN THE GENES

In classical genetics the gene was thought to be a unit or block of material within which no crossing-over occurred. Then, in the late 1940's—much to the surprise of most geneticists—crossing-over between two eye-color mutants, apricot and eosin, was demonstrated. These had been regarded as alleles.

It is now clear that recombination occurs within the gene—it is typical rather than exceptional. As will be discussed in Chapters 12-14, this has been one of the means by which the fine structure of the gene has been revealed.

MOLECULAR DETAILS ARE LARGELY UNKNOWN

You might expect that the new detailed knowledge of DNA and chromosome structure would have provided an understanding of synapsis, chromosome separation, and crossing-over. How the homologous chromosomes find each other and pair precisely during the meiotic prophase is still not understood. There is a characteristic structure, the *synaptinemal complex,* that is invariably associated with synapsis and crossing-over. It is not found in meiotic cells of Drosophila males where there is no crossing-over. But how it works is not understood. Nor are the mechanics of centromere activity and mitosis.

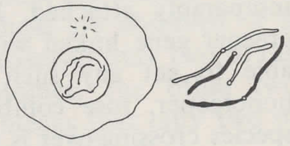
Of course there is speculation and plausible models have been suggested. These are based mainly on experiments with bacteria and viruses. However, there are possibilities for understanding the process in higher organisms, such as yeast and Drosophila, by utilizing mutants that disrupt the meiotic process and recombination; a number of such mutants are known. We shall return to a further discussion in Chapter 16.

MEIOSIS

Entire Cell **Chromosomes**

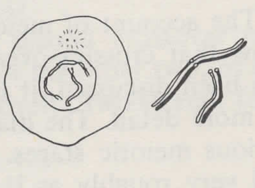
Prophase I. (Leptotene Stage)

The chromosome threads are uncoiled and lying at random in the cell nucleus. The centrosome, above the nucleus, has not yet divided. This is a hypothetical species with four chromosomes (two pairs). Actually each chromosome is split into two chromatids, but usually this is not visible.



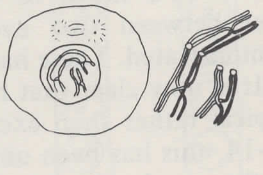
Prophase I. (Zygotene Stage)

Each paternal chromosome (black) has paired with the homologous maternal chromosome. The threads are still uncoiled and are to some extent wound around each other. The pairing is exact so that corresponding genes lie opposite each other. The centrosome has divided into two which are moving toward opposite sides of the nucleus.



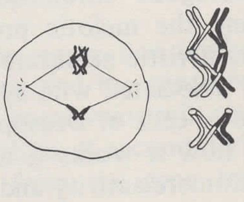
Prophase I. (Diplotene Stage)

Each of the chromosome threads has become visibly double so that the chromosomes are associated in groups of four, called tetrads. Several crossovers have occurred, each involving two of the four strands. The X-configuration that is seen as a consequence of a previous crossover exchange is called a **chaisma** (plural **chiasmata**).



Metaphase I.

The chromosome threads have coiled so that the chromosomes appear much shorter and thicker. The centrosomes are now at opposite ends of the cell and the chromosomes are lined up in a plane at the center. The nuclear membrane has disappeared.



Anaphase I.
Each tetrad
dyads which a
site ends of th
overs each dya
paternal and
material. It is
which directio

Anaphase II.
Each dyad ha
two compone
ing to each p
the stages bet
and the seco
much shorten
uncoiling of th

The four c
meiosis. In a
develop into a
one will form
the remaining
polar bodies.
cells in a ma
grains. In the
egg three cells
bodies in fer
remaining cell
egg after un
divisions. In
same process
four cells dev
which the hap
cycle develop.

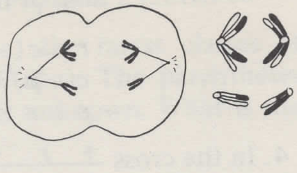
1. A rat has
were mapp

2. In Drosoph
locus 57.0
62.0. How
among 100

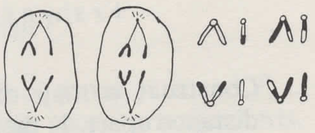
+

B

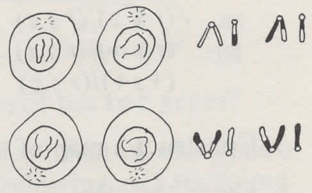
Anaphase I.
Each tetrad has separated into two dyads which are moving toward opposite ends of the cell. Because of crossovers each dyad contains a mixture of paternal and maternal chromosome material. It is a matter of chance in which direction a given dyad goes.



Anaphase II.
Each dyad has now separated into its two components, one of which is going to each pole of the cell. Usually the stages between the first anaphase and the second anaphase are very much shortened, with no appreciable uncoiling of the chromosomes.



The four cells that result from meiosis. In a male animal each will develop into a sperm cell. In a female one will form the functional egg and the remaining three will degenerate as polar bodies. In higher plants the four cells in a male flower form pollen grains. In the formation of the plant egg three cells degenerate (as the polar bodies in female animals) and the remaining cell becomes the functional egg after undergoing three mitotic divisions. In the lower plants, the same process occurs and usually the four cells develop into spores from which the haploid stages of the plant cycle develop.



PROBLEMS

1. A rat has 42 chromosomes in a somatic cell. If all the genes were mapped, how many linkage groups would there be?
2. In Drosophila the dominant sex-linked gene Bar eye (B) is at locus 57.0 on the map. The recessive carnation eye (car) is at 62.0. How many of the various types would be expected among 1000 offspring of the mating?

$$\frac{+ \quad car}{B \quad +}\ \text{♀} \quad \text{by} \quad \frac{+ \quad car}{}\ \text{♂}$$

3. From the following distances between individual pairs of genes construct a map of the chromosome.

$r - c$ 10 units $\quad$ $s - c$ 18 units
$c - p$ 13 units $\quad$ $s - r$ 8 units
$p - r$ 3 units

4. In the cross $\dfrac{+\ r\ +}{w\ +\ s} \times \dfrac{W\ r\ s}{w\ r\ s}$ the following offspring were obtained:

$+r+/wrs$	359	$w++/wrs$	92
$wr+/wrs$	47	$w+s/wrs$	351
wrs/wrs	4	$+++/wrs$	6
$+rs/wrs$	98	$++s/wrs$	43
			1000

Construct a map of the three genes showing their relative distance apart. Is there any interference?

5. In a testcross,

$$\frac{(+++)}{(rsz)} \times \frac{(rsz)}{(rsz)}$$

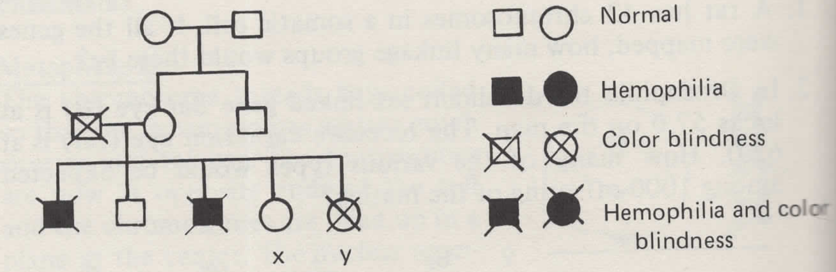
where the parentheses indicate that the order of the genes is unknown, the following progeny were obtained:

$(r+z)/(rsz)$	23	$(+sz)/(rsz)$	75
$(rs\ z)/(rsz)$	396	$(r++)/(rsz)$	73
$(rs+)/(rsz)$	1	$(+++)/(rsz)$	406
$(+s+)/(rsz)$	25	$(++z)/(rsz)$	1
			1000

Construct a map of these factors.

6. From the data in problems 4 and 5 construct a map for the four factors w, r, s, and z.

7. In man the genes for color blindness and hemophilia are both on the X chromosome with about 10% crossing-over. They are an example of linkage of a pathological gene to a relatively harmless one such as could be used for genetic prognosis. The pedigree below shows some of the persons in a more extensive pedigree published by Hoogvliet (*Genetica* 23:94. 1942).



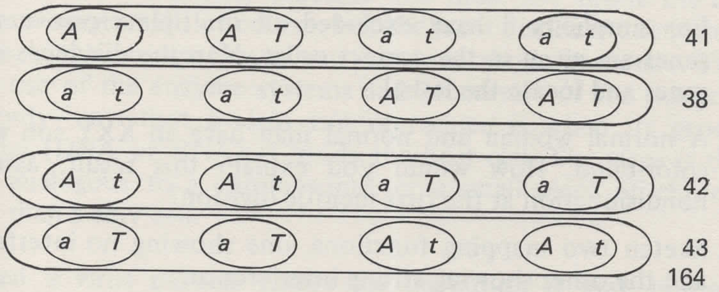
What advice could be given to the women, x and y, as to the likelihood of their having sons with hemophilia?

8. In a cross $(abc)/(ABC) \times (abc)/(abc)$ the rarest classes of progeny were $(abC)/(abc)$ and $(ABc)/(abc)$. The parentheses indicate that the order of the genes is unknown. What is the correct sequence of the three genes?

9. In Drosophila, crossing-over occurs in the female but not in the male. Genes A and B are 10 crossover units apart. What genotypes and in what proportions are expected from the cross

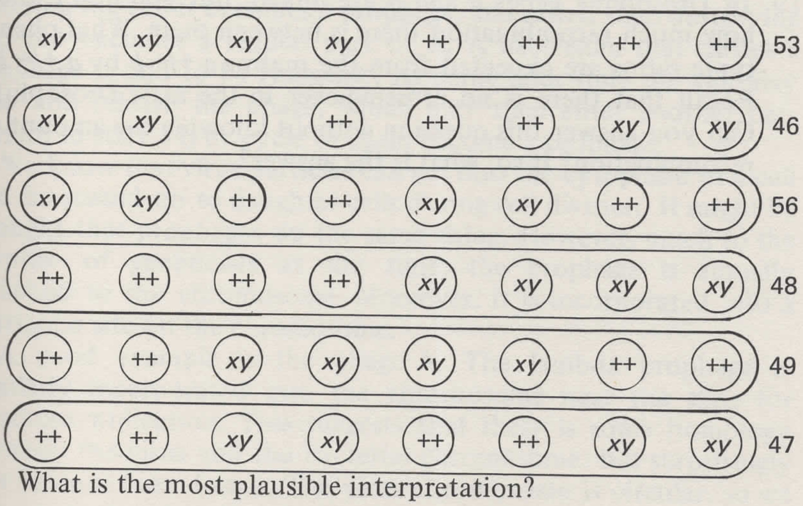
$$\frac{A\ B}{a\ b}\ \text{female} \times \frac{A\ B}{a\ b}\ \text{male?}$$

10. A Neurospora cross involving two pairs of genes, A, a and T, t, produced 164 asci with the following spore arrangements. (Only one member of each pair of spores is shown, since both members of a pair are identical.)


A T A T a t a t 41
a t a t A T A T 38
A t A t a T a T 42
a T a T A t A t 43
164

What can you say regarding the location of the two genes?

11. Suppose you examine ascospores from Neurospora heterozygous for two factors, x and y. The following asci are found:


xy xy xy xy ++ ++ ++ ++ 53
xy xy ++ ++ ++ ++ xy xy 46
xy xy ++ ++ xy xy ++ ++ 56
++ ++ ++ ++ xy xy xy xy 48
++ ++ xy xy xy xy ++ ++ 49
++ ++ xy xy ++ ++ xy xy 47

What is the most plausible interpretation?

12. If the average number of chiasmata per tetrad is two, what is the length of the genetic map of this chromosome?

13. Does it bother you that a chromosome can have a map length of greater than 100 units when each unit means 1% crossing-over?

14. A certain kind of fungus has the four meiotic products in a single ascus, but they are in no particular order. Could you map the distance of a gene from the centromere in this species?

15. A female Drosophila has X chromosomes of the constitution $\frac{+\,+\,+\,+\,+\,+}{a\,b\,c\,d\,e\,f}$.

A recessive lethal is located somewhere in the upper chromosome, but you don't know where. The phenotypes of 1000 male progeny of this female were:

a b c d e f	750	*d e f*	70
a b c d e	60	*c d e f*	40
a b c d	20	*b c d e f*	30
a b c	30	Total	1000

For simplicity I have excluded all multiple crossovers. The genes are given in the correct order. Map the distances of the genes and locate the lethal.

16. A normal woman and normal man have an XXY son who is color-blind. How would you explain this result, assuming nondisjunction at the first meiotic division?

17. Sketch two mapping functions, one showing no interference and the other showing strong interference.

18. In Drosophila, how could you distinguish between genes that are very far apart on the same chromosome and genes on independent chromosomes?

19. In Drosophila genes *a* and *b* are linked, but you don't know how much recombination there is between them. What phenotypic ratios are expected from the mating *a* +/+ *b* by *a* +/+ *b*? Recall that there is no crossing-over in the male Drosophila. Can you answer this question without knowing the amount of recombination? If so, what is the answer?

9

BACTERIAL AND VIRAL GENETICS

VIRUSES

In Chapter 7 we noted that bacteria and viruses have a sexual process similar in many ways to that in higher organisms. It is also clear that they have genetic material organized in linear fashion as are the chromosomes of eukaryotes. The chromosome, however, is simply a double helix of DNA.

The viruses are strictly parasitic and must live inside the cell. They are very small, of course; but the most distinctive feature of viruses is their incapacity to carry out metabolic processes. They make use of the enzyme systems and precursors of the host cell. The only way that a virus can reproduce is when its genetic material (DNA or RNA) is inside the host cell. The virus is thus more analogous to a chromosome, or a small piece of chromosome, than a tiny cell.

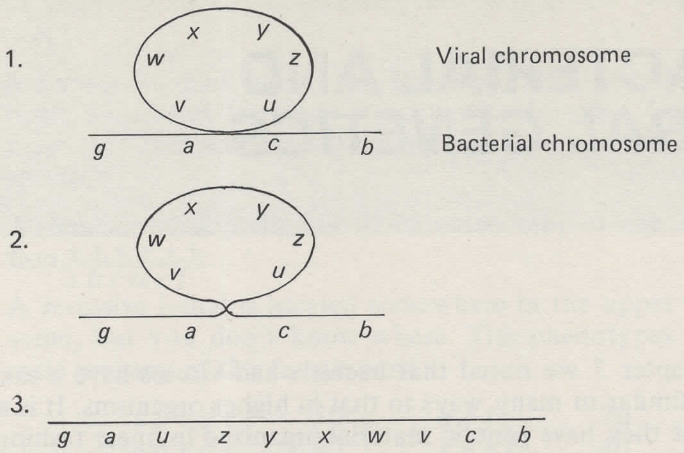
LYSOGENY

When a virus particle enters the bacterium it may multiply within the cell until the cell finally lyses and releases the new virus particles, as was described in Chapter 7. This is called the **lytic** cycle.

In some instances a virus may behave quite differently. After entering the cell it becomes a **prophage**, and rather than destroying the cell becomes a regular part of it. A bacterium that carries a prophage is said to be **lysogenic**. At some later time the cell may lyse and release the phage, which may then enter another bacterium to start a lytic cycle or again become a prophage.

We know that virus particles can get into the cytoplasm of a cell and be passed on to daughter cells during cell division. It might be thought that prophages do the same thing. However, much to the surprise of geneticists at the time, the prophage is actually attached to the chromosome. Moreover, it is incorporated into a particular site on the chromosome.

A good example is the phage λ. The lambda prophage is regularly incorporated into the chromosome near the gene for galactose utilization. This suggests that there is some homology between the virus and the bacterial chromosome, but surprisingly this has not been found. The virus chromosome is circular, so we can easily imagine the process as taking place by synapsis and crossing-over. If *u, v, w, x, y,* and *z* are virus genes, and *g* and *b* stand for the galactose and biotin loci in the bacterial chromosome, we can diagram the process as follows:



Thus the virus is incorporated as an integral part of the bacterial chromosome. The release of the virus can be imagined as the reverse process, 3 → 2 → 1.

There are various other known viruses of this sort, each associated with a specific chromosomal site.

Sometimes the virus causes a change in the bacterial phenotype and its presence can be determined this way. An example is found in the diphtheria organism, where the virus causes the production of a toxin by those cells carrying it. More often, the virus causes no detectable change in the bacterium, and its presence can only be determined when the virus is released in the presence of another strain of bacteria on which the virus is lytic. For some reason, ultraviolet treatment often causes the prophage to be released, so the way in which a lysogenic bacterium is identified is to treat a culture with ultraviolet and then add it to a strain of bacteria in which the phage is lytic.

An entity such as this, that is sometimes on the chromosome and sometimes not, has been called an **episome.**

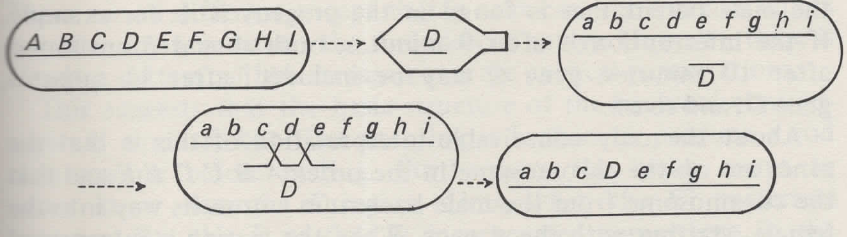
TRANSDUCTION

Sometimes a virus, when it leaves one bacterial cell and infects another, causes a change in the genetic makeup of the second cell. For example, a virus that leaves a bacterium of genotype *A* and enters one of genotype *a,* converts the second one into *A,* as if the virus carried the gene from one bacterium to the other. This phenomenon, first discovered by Zinder and Lederberg, has been called **transduction.**

In the process of transduction, only a small piece of chromosome, carrying at most a few genes, is transferred. If a virus comes from a bacterium of genotype *A B C D E F G H I* and enters a bacterial cell of type *a b c d e f g h i* the types that may be produced by transduction are *A b c d e f g h i, a B c d e f g h i, a b C d e f g h i,* etc., occasionally types as *A B c d e f g h i, a B C d e f g h i,* etc., but not such types as *a b C D E F G h i* or *a b C d e f G*

$\underline{h\ i}$. If two genes are transduced simultaneously they are always closely linked.

Since the bacterium after transduction is haploid, there must be a substitution of the transduced material for the original genes in the recipient bacterial cell. Furthermore the material is inserted in the right place in the chromosome. Perhaps it is by a process analogous to synapsis and crossing-over, and could be diagrammed this way:



The diagrams above indicate one possibility that is consistent with the data; since none of these processes can actually be seen, it is not possible to say whether or not this is the actual mechanism.

Simultaneous transduction of two or more genes can be used as a basis for construction of a chromosome map. For example, if gene combinations *AB, BC,* and *CD* are transduced, but not combinations such as *AC* and *AD,* the gene order is inferred to be *ABCD.* In this way it is possible to construct a map for bacteria that can be transduced, but which do not have an ordinary sexual process.

The process that we have been discussing is **generalized transduction.** The gene that is transduced may come from any part of the donor bacterium, and is carried by a lytic virus.

There is also **specialized transduction** in which the agent is a temperate virus, behaving as an episome. A prophage, such as λ (which occupies a specific site in the bacterial chromosome), can carry some adjacent genes with it when it leaves the chromosome. Then, when it enters the host cell and crosses-over with the host chromosome, these genes are incorporated. The process is called specialized because the genes carried are only those that are closely linked to the normal position of the transducing virus.

MULTIPLE RESISTANCE TRANSFER FACTORS

Some bacteria are resistant to many kinds of antibiotics and hence constitute a severe medical problem. They carry one or more virus-like particles that have incorporated various genes for resistance. The particle has the necessary DNA for replication and for cell infection and, presumably by some sort of recombination process, has picked up genes from some source that confer resistance. Such viruses are transferred from cell to cell by contact, so the population of resistant bacteria increases enormously. The more different kinds of resistance the virus can incorporate, the greater are its chances of increasing. Furthermore, the factor can be transferred to another species, so resistance factors from harmless bacteria such as *E. coli* get into pathogenic bacteria such as Salmonella and Shigella. Indiscriminate use of antibiotics hastens the spread of such resistant strains.

PARTIAL TRANSFER OF GENES IN MATING BACTERIA

Two French geneticists, F. Jacob and E. Wollman, discovered that, if they separate mating pairs of bacteria by beating them up in a blender, the genes transferred from the male to the female differ according to the time when the mating was interrupted. If mating is interrupted after about 7 minutes, the only gene from the male parent that is found in the progeny is *A*, for example. If the interruption is after 9 minutes, both *A* and *B* are found; after 10 minutes, gene *C* may be included; after 11 minutes, gene *D*; and so on.

About the only conceivable interpretation of this is that the genes are on the chromosome in the order *A B C D E F* and that the chromosome from the male bacterium worms its way into the female, starting with the *A* gene. When the mating is interrupted after 10 minutes, for example, the genes *A, B,* and *C* have entered the female cell and in tearing the cells apart the chromosome is broken between *C* and *D.* Thus, Jacob and Wollman have constructed a chromosome map in which the map distance corresponds to time units; on this scale genes *B* and *C* are 1 minute apart. The gene order inferred in this fashion agrees with that determined by ordinary gene mapping by crossing-over.

This shows that in bacteria there can be various intermediates between full sexuality, in which a complete haploid set of genes is transferred, and transduction, in which only a tiny chromosome fragment is transferred.

SEX DETERMINATION IN BACTERIA AND ENDLESS THE CHROMOSOME

Early in the history of bacterial genetics it was discovered that there are two mating types and that the presence of these types depends on the presence or absence of a fertility factor *F.* The male cells were designated F^+ and the female F^-. The F^+ factor has the peculiar property of being readily transferred from cell to cell on contact, so that F^- cells that are in the same culture with F^+ are converted into F^+. F^- cells are simply cells that lack the *F* factor.

The *F* factor is an episome. When the particle is free in the cell the bacterium is only potentially capable of donating a chromosome, or partial chromosome, to a female cell. The particle has to be integrated into the chromosome for this to occur. Bacteria in which the *F* factor is part of the chromosome are called Hfr (for high frequency of recombination).

When mating occurs and the chromosome from the male enters the female cell the *F* region is always the last to enter, and therefore it frequently gets left out. The conversion of an F^+ strain into Hfr, by attachment of the free *F* particle to the chromosome probably takes place by a process similar to that described above for the incorporation of λ phage into the chromosome. The *F* particle is circular like λ.

As stated before, the gene sequence can be mapped by the time of entry into the female cell. Each Hfr strain gives a consistent map. However, when different Hfr's are studied the gene orders are different. Consider the following five hypothetical Hfr strains, with gene orders determined by interrupted mating experiments:

Hfr 1. ← *a b c d e F*
Hfr 2. ← *c d e a b F*
Hfr 3. ← *d e a b c F*
Hfr 4. ← *d c b a e F*
Hfr 5. ← *e a b c d F*

These are all different except that in each case the *F* region enters last (the arrow indicates the direction of entry). However, there is clearly a pattern; all the arrangements are cyclical permutations.

This suggests that the basic structure of the bacterial chromosome is circular, but that the ring is broken to produce a rod chromosome before mating. Furthermore the break seems to occur at or near the *F* region. Thus, when the F^+ bacterium becomes Hfr the *F* particle is incorporated somewhere in the chromosome. This determines the gene entry order for this particular strain. Another strain in which the *F* factor is incorporated in a different region has another order.

There is some evidence that the chromosome enters the female cell as it is being duplicated. One possibility is that duplication starts at the *F* region and that the new strand is broken at this point. As duplication proceeds the new strand peels off and enters the female cell.

At the time that these genetic studies were done the idea of a circular chromosome came as a surprise. Now, as has been mentioned various times before, there is direct evidence from electron microscopy and autoradiography that the chromosome is actually a closed loop.

In some bacteria the *F* particle (designated *F′*) differs in that all Hfr strains derived from it have the particle incorporated into the chromosome in the same place. The most plausible explanation is that *F′* includes a small bit of the bacterial chromosome. This then could synapse with the homologous region of bacterial chromosome and insure that the *F′* particle is incorporated at the same position each time.

F′ strains have been very useful in genetic studies. Ordinarily the bacterium is haploid so dominance relations cannot be studied. With one or more *F′* particles in the cell the organism is diploid (or polyploid) for those genes carried by the particles. Local diploidy obtained in this way has been useful not only in the study of dominance, but also in throwing some light on the mechanisms of regulation of gene action, as will be discussed in Chapter 18.

TRANSFORMATION

Long ago, in the 1920's, it was discovered that some peculiar transformations could be carried out with the Pneumococcus. Virulent bacteria have a capsule surrounding the cell. It was discovered that nonvirulent cells without capsules, when injected into a mouse along with dead encapsulated cells, were able to kill the mouse, and virulent cells were recovered. Later the experiments were repeated *in vitro.* Still later it was shown that the transformation from nonvirulent to virulent could be accomplished with only an extract from the encapsulated cells. Furthermore, when an extract from a virulent strain was added to a nonvirulent strain the strain became virulent and had the same antigenic properties as the strain from which the extract was made. Only a tiny fraction of the cells were thus transformed, in either the *in vivo* or *in vitro* experiments, but it is possible to identify a tiny minority of virulent cells.

Over a period of years the extract was purified until it was finally discovered by Avery, MacLeod, and McCarty that the transforming material was virtually pure DNA. It was also shown that various genetic traits, and not just capsular properties, could be transformed this way. During the 1940's this provided some of the strongest evidence for the growing belief that DNA is the genetic material.

It was established early that the changes brought about by the transformation were permanently incorporated into the heredity, for the transformed property "bred true" in later generations. Now, in view of our knowledge of lysogeny and transduction, it does not seem so surprising that DNA from one strain can be incorporated into the chromosome of another. But in the 1940's this was a very puzzling finding. An early and quite prophetic statement was made by H. J. Muller in a famous lecture published in 1947. He suggested that in transformation "parts of chromosomes might have penetrated the capsuleless bacteria and in part at least have taken root there, perhaps after having undergone a kind of crossing-over with the chromosomes of the host."

There is the important question of whether transformation can occur only in bacteria, or whether something of the same kind occurs in higher forms. So far there is no convincing evidence for transformation in higher forms. However there are some little-understood genetic changes induced by DNA in Drosophila. Also, it has been possible to change human cells in culture from drug susceptible to drug resistant by adding DNA extracts from resistant cells. However, since further genetic tests are not possible in this system it is not possible to say whether the DNA has actually been incorporated into the chromosome.

Sometimes, as with transduction, two closely linked genes will be carried together during transformation. In this way it is possible to determine gene sequence just as with transduction. In fact, transformation and transduction appear to be fundamentally the same; in one case (transformation) the DNA is transferred from one cell to another directly, in the other it is carried by a virus. In either case it is incorporated into the bacterial chromosome and becomes a part of its heredity.

LINKING PIECES OF DNA TOGETHER

Although the general method by which broken chromosomes unite is not known, there are some mechanisms that are known to work experimentally and which are very suggestive as to how the actual process may occur.

One class of enzymes, known as **restriction enzymes**, are very specific as to the DNA regions that they attack. Each enzyme recognizes a specific short DNA sequence. Such regions are usually, if not always, reverse repeats, such as:

$$\begin{array}{c}\overline{A\ G\ T\ A\ C\ T}\\ \underline{T\ C\ A\ T\ G\ A}\end{array}$$

The enzyme cuts the DNA so as to leave "sticky" ends of the type:

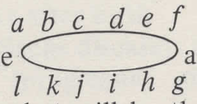
```
_____________________________                       _________
                      A G T A C T
                                     and
___________                            T C A T G A _____________
```

Such ends can grow back together by forming hydrogen bonds by the usual Watson-Crick rules. But they can also join to any other sticky ends with the same sequence. The DNA molecule is so long that short sequences of the same kind occur at several places along the chromosome, so if a restriction enzyme cuts all these they can join in new ways to produce new gene orders.

In the laboratory this has been an exceedingly useful trick. By the use of a restriction enzyme on DNA of different kinds, homologous sticky ends can be produced. In this way it has been possible to link together DNA from different species, such as *E. coli* and Drosophila.

The possibility of such use in genetic engineering at once suggests itself. Could we join the normal gene for antihemophilic globulin to a virus that is capable of entering the human cell, and thereby hope to cure the blood disease? At the moment this is far-fetched, but molecular knowledge is growing at an enormous rate. It also suggests caution, for there is also the possibility of accidentally making tumor-inducing episomes or new kinds of drug resistance factors.

PROBLEMS

1. When a strain of genotype *A C N R X*, where the gene order is unknown, was used as a source of DNA used to transform a strain of genotype *a c n r x*, the following types were found: *A c n R x*, *a c N r X*, *a C n R x*, and *A c n r X*, in addition to those such as *a C n r x* where only a single gene was transformed. What is the gene order?
2. Given the circular chromosome  and with the *F* factor inserted between *i* and *j*, what will be the order of gene entry during conjugation?
3. During the lytic cycle the viral chromosome may undergo repeated "matings" (i.e., synapsis and gene exchange). Suggest a way in which this might be proven.

4. Sometimes there is "abortive" transduction; the gene that is transferred, instead of attaching to the chromosome, remains in the cytoplasm. It can function, but cannot reproduce. Describe the inheritance of the trait determined by this gene in subsequent cell divisions.
5. Why are lysogeny and specialized transduction easier with a circular chromosome than a rod-shaped chromosome?
6. What is the possible significance of the reverse repeat in the sequence recognized by restriction enzymes?
7. Where do you suppose the genes for antibiotic resistance in the multiple-resistance transfer factors came from?

GENE INTERACTION AND THE EFFECT OF THE ENVIRONMENT

10

GENES MAY AFFECT IMPORTANT OR TRIVIAL PROCESSES

Some genes affect extremely minor characters. Genes have been studied in Drosophilia which affect slightly the shape of certain bristles, or change the number of hairs on the wing. At the other extreme are genes which cause gross pathological abnormalities such as dwarfism, complete mental breakdown, or extremely scaly skin. In plants the failure of a certain gene to act properly may disrupt the entire photosynthetic process, the basic reaction on which all life depends. There are many genes which are essential for life. When such a gene is missing or is replaced by a mutant allele the animal or plant dies. Such mutant alleles form the large class of **lethal** genes. Between these extremes are the great bulk of gene differences which the geneticist studies, genes affecting hair color, skin color, intelligence, yield in farm crops, fertility and rate of growth in livestock, etc. Before a gene effect can be studied effectively by a geneticist (1) the gene must exist in two or more forms, (2) its effect must somehow be observable.

GENE-DETERMINED CHARACTERS MAY APPEAR AT ANY TIME

A commonly held, but false, belief is that hereditary characters are always congenital (i.e., present at the time of birth). Hereditary characters may develop at any stage of development. Each gene has its characteristic time for action and this may be any time from the zygote immediately after fertilization to the time of death. In corn, Drosophila, and the house mouse, genes have been found affecting almost every developmental process at various times and the same would be true for other species if they were as well studied. As examples of gene-determined characters appearing at different stages of development the following list of human hereditary traits is included:

1. polydactyly (extra fingers and toes): early embryonic stage
2. eye pigment: a few days or weeks after birth
3. infantile amaurotic idiocy (Tay-Sachs disease): about one year
4. X-linked muscular dystrophy (progressive muscle deterioration): 10 to 15 years
5. hereditary baldness: 25-50 years
6. Huntington's chorea (progressive mental and nervous deterioration): 30-50 years

MANY GENES AFFECT THE SAME CHARACTER, AND VICE VERSA

When a gene is studied carefully, it is usually found that it has several effects. For example, the gene for white eye in Drosophila also affects the color of the testes and the shape of the sperm receptacles. It has long been known that blue-eyed white cats are deaf. Mink breeders are familiar with the fact that animals with the colmira color gene are deep sleepers and that certain pastel-colored types hold their heads in a peculiar way. The most extreme example of a gene affecting many characters is the gene "polymorph" in Drosophila: it affects the eye color, body proportions, wing size, wing vein arrangement, body hairs, size and arrangement of bristles, shape of testes and ovaries, viability, rate of growth, and fertility. In man the gene for Eddowe's syndrome results in the normally white sclerotic coat of the eye being blue, and brittle bones. The gene for epiloia results in skin tumors, moles and birthmarks, brain growths, feeblemindedness, and sometimes, epileptic seizures. A gene that has several superficially unrelated effects is said to be **pleiotropic** (noun: pleiotropy).

Conversely, just as one gene often affects many characters so do many genes affect a single character. As an example some 20 gene loci are concerned with the pigment of the eye in the fruit fly, and more than this are involved in the color of a corn plant. One human condition, epidermolysis bullosa, or blistering skin, is caused by at least four different genes. The same is true of muscular dystrophy, and there are numerous other examples.

Many quantitative traits, such as height and weight, are determined by a large number of genes with individual effects usually too inconspicuous to be identified. Such traits are almost always affected by the environment as well, and it is often difficult—particularly in the human species—to separate genetic from environmental influences when complex inheritance is involved.

WHAT IS "NORMAL" OR "WILD TYPE"?

From what has been said it is clear that there is no such thing as a gene for the eye, another for the ear, and so on. Each part of the body is influenced by many genes, and each gene produces effects that may ramify to all parts of the body.

When all the genes function normally, the individual develops normally. When a gene is missing or has mutated to an abnormal form, the normal result is somehow modified. The modification may be great or small, important or unimportant. What the geneticist knows about the normal processes he learns by inference from what goes wrong when a normal gene is replaced by a mutant. The fact that normal red-eye pigment is not produced by a fly whose *W* genes have been replaced by mutant *w* genes tells us that the *W* gene had something to do with the eye pigment. But it doesn't tell us that this is *the* gene that makes eye pigment; probably dozens of other genes are also involved in the same process.

POLYMORPHISM

Although most of us have an idea of what a "normal" or "wild-type" Drosophila is, this does not mean of course that all "normal" flies are alike. To the expert there are many differences. This is even clearer if we look at species close to ourselves. Sometimes the difference is conspicuous, as between the black and cinnamon phases of the American bear, both normal. This is an example of **polymorphism**, two or more forms normally existing in the same species. The human species is strikingly polymorphic with our variety of hair and eye colors, to say nothing of shapes and sizes.

The word *polymorphism* has been extended to include invisible and behavioral traits as well. Examples are blood groups (to be discussed in Chapter 19) and a number of serum proteins. The picture that emerges is that, although the phenotype is relatively constant, there is an enormous genetic diversity; many genotypes produce very similar phenotypes. Such cryptic variability is the raw material for future evolution.

VARIATION IN GENE EXPRESSION

Some genes are very constant in their expression, others are extremely variable and there are all grades in between. Sometimes the variability is due to variable environmental factors or other genetic factors that can be controlled; in other cases no such factors can be found. In Drosophila most of the genes for eye color produce extremely constant effects though usually the color darkens somewhat with age. On the other hand, the gene "Lobe," which affects the size and shape of the eye, may result in an eye only slightly different from normal or it may be reduced to pin-point size. In man the gene for polydactyly (extra fingers and toes) may cause only a slight knob of an extra finger or two or more well-formed extra digits. Genes whose expression varies are said to be of variable **expressivity**.

PENETRANCE

Another form of variation in gene expression is variable **penetrance**. Sometimes the character usually associated with a certain gene fails to appear at all. A gene for which this is the case is said to have reduced penetrance. Penetrance is defined as the proportion of cases known to possess the proper genotype in which the character actually appears. A gene which is always expressed has a penetrance of 100%; many have a lower value. An example of a gene with incomplete penetrance, in this case 90%, is the gene, *i*, causing an interrupted wing vein. Of flies with genotype *i/i* 90% have an interrupted wing vein, the other 10% are normal. That the normal flies have the same genetic constitution as the others is shown by the fact that they, too, produce 90% offspring with interrupted veins.

DELAYED GENE EXPRESSION

In some snails the direction of coiling is hereditary but is inherited in what at first appears to be a very complex fashion. In the hybrid between a right- and left-coiled pure strain the direction of coiling is always the same as that of the mother. When these F_1 hybrids are self-fertilized the F_2 are all right-coiled. Then in the F_3 there is a 3:1 ratio of right to left. The inheritance is simply explained by noting that the character is one generation late in developing. The direction of coiling is determined, not by the snail's own genotype, but by the genotype of its mother. There is a delay in the expression of the genes until the next generation.

The most extreme example of lag in gene expression is the mutant, grandchildless, in Drosophila. Females homozygous for this gene are fertile but their offspring are sterile. The factor doesn't have any effect on males or their offspring.

ENVIRONMENTAL EFFECTS ON GENE EXPRESSION

Often the manner of expression or the penetrance of a gene is directly related to controllable changes in the environment. One of the easiest environmental factors to control in genetic experiments on laboratory organisms is temperature. There are many genes known in Drosophila whose penetrance varies from almost 0 to almost 100% as the temperature is changed. In others the manner of expression changes with temperature. For example, the character balloon wing (wings with blisters) is much more extreme at 19° than at 25° centigrade.

Another example of a gene depending on temperature is the gene c^h for himalayan pattern in rabbits. Rabbits of genotype c^h/c^h are white with black at the tips of the feet, ears, nose and tail. This is due to the fact that the enzyme responsible for the dark pigment is inactivated at normal body temperatures so that only those parts which are normally cooler develop color. This has been tested experimentally by placing an ice pack on part of the rabbit. The hairs that grow in this area are dark.

In corn there is a gene, sun red, whose expression is related to the amount of sunlight. Sun-red corn plants are red in those parts where the sun strikes, green in shaded parts.

A human disease is also affected by sunlight. The recessive condition, xeroderma pigmentosum, produces a skin freckling. But if the person is exposed to much sunlight, skin tumors develop. In this case the molecular explanation is known. What is inherited is the inability to repair DNA damaged by ultraviolet light.

Another environmental factor, in this case the internal environment, is a hormone. A good example is early baldness. Since the affected individuals are essentially all males, it might be suspected that the inheritance is X- or Y- linked. However, the trait is due to a dominant autosomal gene; but the gene is expressed only in the presence of the male hormone.

Sometimes the dominance is reversed in the two sexes. There is evidence in Ayrshire cattle that mahogany color is dominant in males while red is dominant in females.

IMPORTANCE OF THE ENVIRONMENT

Enough examples have been given to show that the type of environment is extremely important in determining the final character that a gene will produce. Normal development of a character requires not only the proper set of genes but also the proper environment. In a course emphasizing the mechanism of heredity it is easy to overlook the importance of the environment in shaping the final character. We shall try to avoid that error.

Sometimes a hereditary character may resemble an environmental one. For example, diabetes may be hereditary or due to injury to the liver or pancreas. In most cases of complex characters, such as health, size, intelligence, disposition, and talents the genetic and environmental factors are so complex and so interwoven as to make it impossible to separate them.

As ordinarily stated, the question—Is heredity or environment more important?—is meaningless. *Both* the proper heredity and the proper environment must be present for the realization of the character. But it is possible to ask whether within the usual range of variability in genes and environment in the population most of the variability in a character can be attributed to gene differences or to environmental differences.

Just because a disease follows a characteristic hereditary pattern doesn't mean that it cannot be influenced by environmental means, or that it is incurable. For example, diabetes and pernicious anemia are diseases with a strong hereditary predisposition—yet both can be successfully treated.

USE OF TWINS FOR STUDYING HEREDITY AND ENVIRONMENT

Twins are of two kinds: (1) one-egg, or identical, or monozygotic; and (2) two-egg, or fraternal, or dizygotic. One-egg twins come from a single fertilized egg and have identical genetic constitutions. Two-egg twins come from two fertilized eggs that happened to be produced at the same time, and are no more alike genetically than ordinary brothers and sisters. In the United States about 1/3 of the twins are one-egg.

It is common knowledge that one-egg twins look strikingly alike. They agree perfectly for eye color, hair color, and other strictly inherited traits. There are also similarities in physiological and behavioral traits. H. Kalmus made the interesting finding that dogs have some difficulty in distinguishing between identical twins by smell. He took dogs trained to identify and retrieve a handkerchief that had been handled by a particular person from among a number of handkerchiefs handled by other persons. The dogs did this unerringly except where identical twins were involved; in this case they got the wrong twin about as often as the right one.

Kallmann reported that when one of a pair of one-egg twins is hospitalized for schizophrenia, his co-twin has a probability of about 85% of developing schizophrenia during his lifetime. The corresponding value for two-egg twins and sibs of the same sex is about 15%. Other investigators have found quite different absolute values, usually lower, but all the studies agree in a much higher concordance for one-egg than for two-egg twins.

The greater similarity in the one-egg twins argues that genetic factors play some role in the causation of schizophrenia, although it could be argued that one-egg twins because of their similarity encounter more similar environments. On the other hand, since the concordance is not 100%, it is clear that genetic factors are not the sole cause. Further support for the influence of genetic factors comes from children who were adopted as infants. Those who developed schizophrenia had about four times as many biological relatives with the disease as those who did not.

To consider a totally different trait, the average difference in life span of one-egg twins in one study was 37 months, whereas it was 78 months for two-egg twins. Again, there is the possibility that the one-egg twins had a more similar environment.

The most definitive information comes from twins that were separated at an early age and reared in different homes. One such study by H. H. Newman yielded the following data on height, weight, and IQ.

Table 10-1

Character	Mean pair difference			
	One-egg (50 pairs)	Two-egg same-sex (52 pairs)	Siblings same-sex (52 pairs)	One-egg reared apart (19 pairs)
Height (cm)	1.7	4.4	4.5	1.8
Weight (lbs)	4.1	10.0	10.4	9.9
Binet IQ	5.9	9.9	9.8	8.2

This type of analysis, even if the data are taken at face value, doesn't permit a quantitative conclusion. We could say that height seems to be more determined by genetic factors than weight, with IQ somewhere between; but the very small number of one-egg twins reared apart keeps us from having any confidence in the statement. That genetic factors play some role in IQ is supported by the fact that in all studies of separated twins, one-egg twins reared apart have been more similar than two-egg twins reared together.

We shall return to the question of quantitative traits later in the book, and develop more quantitative methods.

GENE INTERACTIONS

When two or more genes act on the same character one gene may mask the effect of the other, in much the same way as a recessive gene effect is masked by that of its dominant allele. This phenomenon of a gene masking the effect of another nonallelic factor is called **epistasis**.

An example of epistasis is found in coat color in mice. Gene C is necessary for any pigment; the genotype c/c is albino. The genotypes B/B and B/b result in black color and b/b in brown, but these genes can only be expressed in the presence of the C factor which is necessary for any color at all.

The various genotypes and their phenotypes are:

$$\begin{array}{ll} C/-\;B/- & \text{black} \\ C/-\;b/b & \text{brown} \\ \left.\begin{array}{l} c/c\;B/- \\ c/c\;b/b \end{array}\right\} & \text{white} \end{array}$$

A cross between two individuals each of genotype $C/c\ B/b$ would produce on the average 9/16 black, 3/16 brown, and 4/16 white offspring. This is a typical 9:3:3:1 ratio with the last two classes added together.

In this instance the c gene is said to be **epistatic** to the B, b pair, since the distinction between B and b is suppressed in the cc genotype. Conversely, the B, b pair are said to be **hypostatic** to the c gene.

Epistasis is also used in a more general sense to denote any instance in which the effect of two pairs of genes is different from the sum of the effects when one pair is considered at a time. This is considered further in Chapter 23.

Another example of gene interaction is found in poultry (as well as many other animals). Both genes C and D are required for pigment production. If either is missing the fowl is white. These two genes produce enzymes that catalyze successive steps in a chemical process leading to melanin pigment, as in this diagram:

$$\text{Raw material} \xrightarrow{\text{gene } C} \text{Colorless intermediate} \xrightarrow{\text{gene } D} \text{Pigment}$$

Hence the various genotypes and their phenotypes are:

$$\begin{array}{ll} C/-\;D/- & \text{colored} \\ \left.\begin{array}{l} c/c\;D/- \\ C/-\;d/d \\ c/c\;\;d/d \end{array}\right\} & \text{white} \end{array}$$

A cross between two parents, each of genotype $C/c\ D/d$, would lead to 9/16 colored and 7/16 white, since the three types $C/-\ d/d$, $c/c\ D/-$, and $c/c\ d/d$ are indistinguishable.

Some genes, known as **modifiers**, have the effect of altering the expression of another gene. An example is the gene, dimorphos, which changes the form of expression of the wing-affecting gene, vestigial. Not only does it alter the phenotype of the homozygous vestigial fly, but changes its temperature dependence. The vestigial gene in the absence of dimorphos results in a wing that is very tiny and hardly looks like a wing. In the presence of dimorphos the wing is much more normal, and at higher temperatures is hardly distinguishable from a wild-type wing. A familiar example of modifying genes is found in spotted patterns in various animals. The basic spotted pattern, as for example in Holstein cattle, is determined by a single basic gene. But the relative amounts of black and white are determined by a series of modifying factors. Some extreme modifiers completely prevent the expression of some other gene. These are known as **suppressors.**

Consideration of genes as producing enzymes provides an explanation of dominance. Since enzymes are needed in only very small amounts for catalyzing chemical reactions, one gene usually produces more than enough to convert all the available raw material into the product. In rodents and chickens, for example, the genotypes C/C and C/c are black while c/c is albino. The gene product governs one of the chemical steps by which the amino acid tyrosine is converted into melanin. One normal C gene produces as much pigment as two, or almost as much. Frequently heterozygous animals can be detected by careful measurements of pigment, or by the rate of pigment formation, or by occasional unpigmented hairs; but superficially, C/C and C/c are identical. The c gene either produces no product, or one that is ineffective as a catalyst for this reaction.

In the next chapter we shall deal more specifically with gene control over enzyme formation.

PROBLEMS

1. In man, hemolytic jaundice is due to a dominant factor with 10% penetrance. A heterozygous man marries a normal woman. What proportion of the children would be expected to have the disease?
2. Baldness is due to an autosomal dominant (B) expressed only in males. A nonbald man marries a woman whose father was nonbald. They have a child that develops baldness. What was the genotype of the woman?
3. How could you distinguish from pedigree studies between a sex-linked recessive factor and a dominant autosomal factor expressed only in males?
4. In Drosophila, the recessive gene sable, s, causes a black body color. Another gene, the recessive suppressor of sable (t), prevents the expression of the sable gene, so the fly is normal gray color. Thus the genotypes $s/s\ T/-$ are sable; all others are normal ($s/s\ t/t$, $S/-\ T/-$, $S/-\ t/t$). What offspring would be expected from the cross $S/s\ T/t \times s/s\ T/t$?
5. In man the gene D is necessary for a normal ear cochlea and gene E is necessary for a normal auditory nerve. In the absence of either of these factors the individual is deaf. Show how two normal parents could produce a deaf child and how two deaf parents could produce a normal child.
6. In man either the homozygous state of the gene a or the heterozygous (or homozygous) state of the gene B results in retinitis pigmentosa, a form of blindness. Only the genotype $A/-\ b/b$ is normal. A blind woman whose parents were both normal marries a man of genotype $A/a\ B/b$. What proportion of the children would be expected to be blind?
7. In chickens B is a dominant sex-linked gene resulting in barred feathers; b or b/b is solid colored. The gene C is necessary for

any color; *c/c* birds are solid white regardless of the presence of the *B* factor. A barred male mated with a colored, nonbarred female produced 3 solid white, 2 barred, and 3 solid colored chicks. What are the theoretical expectation of chicks in this cross?

8. In chickens the genotype *C/ - D/ - i/i* is colored, all other combinations being white. Leghorns are *C/C D/D I/I,* white Wyandottes are *c/c D/D i/i,* and white Silkies are *C/C d/d i/i.* What offspring would be expected from:
    a. a white Silkie crossed with a white Wyandotte?
    b. a white Leghorn crossed with a white Silkie?
    c. A Leghorn-Silkie hybrid backcrossed to a Silkie?
    d. A Leghorn-Silkie hybrid crossed with a Wyandotte?

9. In Drosophila two genes, the dominant sex-linked *V* and the autosomal dominant *Cn,* are necessary for the development of the brown pigment in the eye. When either is missing the pigment doesn't develop. The autosomal gene *Bw* is necessary for the scarlet pigment. When both pigments are present the eye is normal dark red, when neither is present the eye is white. *Cn* and *Bw* are linked with about 40% crossing-over. What proportion of the offspring of the cross

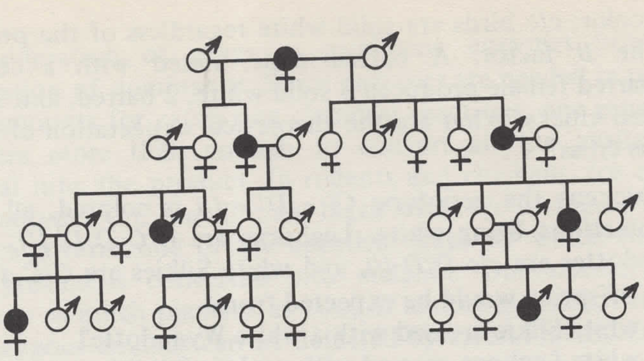
$$\frac{V}{v}\quad\frac{Cn\qquad Bw}{cn\qquad bw}\ \text{female by}\ \frac{v}{\quad}\quad\frac{cn\qquad bw}{cn\qquad bw}\ \text{male}$$

would be expected to have white eyes?

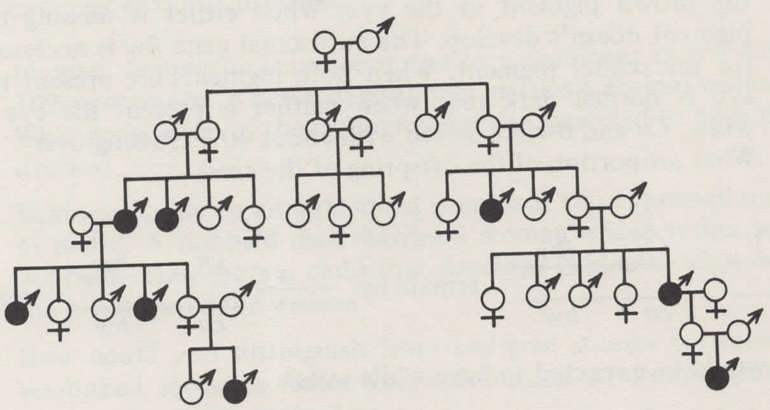
10. A dairy farmer and a geneticist were looking out the farmer's living room window and saw a mahogany-colored Ayrshire cow with a newly born red calf. The farmer wanted to know the sex of the calf, and the geneticist told him that it was obvious from the color which sex the calf was. He explained that in Ayrshires the genotype *AA* is mahogany and *aa* is red; but the genotype *Aa* is mahogany in males and red in females. What sex did he say the calf was?

11. In corn all three dominant factors *A, C,* and *R,* are necessary for colored seeds. Genotypes *A/ - C/ - R/ -* are colored; all others are colorless. A colored plant is crossed with three tester plants with the following results.
    With *a/a c/c R/R* it produced 50% colored seeds.
    With *a/a C/C r/r* it produced 25% colored seeds.
    With *A/A c/c r/r* it produced 50% colored seeds.
    What is the plant's genotype?

12. Here is a pedigree of a family showing a rare disease. Diseased individuals are shown in solid circles, as usual. Is this more likely to be due to a dominant factor with reduced penetrance or to a recessive factor? Explain why.



13. Is this pedigree, also of a rare disease, more likely to be due to a dominant autosomal factor that is expressed only in males or to a sex-linked recessive? Explain why.



14. Siamese twins are usually less similar than separate identical twins. Suggest a reason.

15. Identical twin brothers marry identical twin sisters. Each pair has a child. Legally these children are cousins; what are they genetically?

16. Why is it meaningless to ask the question: which is more important, heredity or environment? What kind of question is capable of being answered?

17. From the following table, which character is most influenced by hereditary differences?

	Percentage of concordance between		
Character	Identical twins	Fraternal twins	Siblings
A	70%	28%	20%
B	95%	90%	40%
C	92%	21%	20%

---

18. Derive the
data the pr
population

Proportion

where P is t

19. Apply this
the followi

U. S.
Japan

20. A Drosopl
childless" i
gene. The
What propo

21. In snails, ri
is fertilized
be expecte
fertilized, w

22. Suggest a h
would be
recessive?

23. A left-coile
coiled proge

18. Derive the following formula for determining from sex-ratio data the proportion of monozygotic and dizygotic twins in the population:

$$\text{Proportion dizygotic} = \frac{\text{Proportion of twins of opposite sex}}{2P(1\text{-}P)}$$

where P is the proportion of males in the twin population.

19. Apply this formula to determine the proportion dizygotic in the following populations:

	Proportion of twins		
	MM	MF	FF
U. S. white	.35	.33	.32
Japanese	.42	.19	.39

20. A Drosophila female heterozygous for the gene "grand-childless" is crossed with a male, also heterozygous for the gene. The female offspring are crossed to unrelated males. What proportion of these females would have sterile offspring?

21. In snails, right coiling (R) is dominant to left (r). An *R/r* snail is fertilized with sperm from an *r/r* snail. What offspring would be expected? If these offspring were allowed to be self-fertilized, what kinds of offspring would be produced?

22. Suggest a hypothesis leading to a 27:37 F_2 ratio. What ratio would be expected if the F_1 is backcrossed to a multiple recessive?

23. A left-coiled snail when self-fertilized produced only right-coiled progeny. What is its genotype?

BIOCHEMICAL GENETICS

11

INTRODUCTION

The subject of biochemical genetics had its beginning with Sir A. E. Garrod, who pointed out in 1902 that the recessive disease alkaptonuria was an inherited defect in metabolism. He wrote a book entitled *Inborn Errors of Metabolism* that is one of the classics of genetics and biochemistry. Modern work on biochemical genetics began with the discovery by Beadle and Tatum of biochemical mutants in Neurospora.

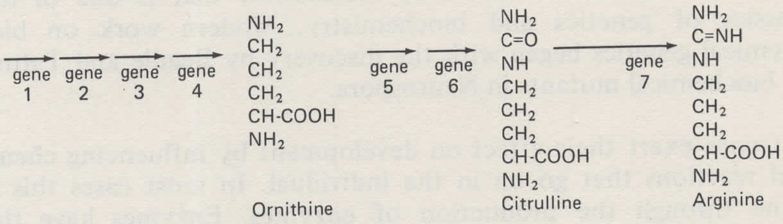
BIOCHEMICAL MUTANTS IN NEUROSPORA

Genes exert their effect on development by influencing chemical reactions that go on in the individual. In most cases this is done through the production of enzymes. Enzymes have the power of speeding up chemical changes without being changed themselves; in other words, they act as chemical catalyzers.

The wild-type strain of Neurospora will grow on a medium containing a carbon source, nitrogen, some salts, and biotin. Since on chemical analysis a Neurospora plant contains all the known vitamins and amino acids, these must be synthesized by the growing plant. In the work initiated by Beadle and Tatum, spores were irradiated, then grown on a complete medium containing all the known vitamins and amino acids (made by grinding up Neurosporas or yeasts, which contain the necessary substances). After growth was established, samples were tested for growth on the minimal medium containing only the carbon source, nitrogen source, and salts plus the vitamin biotin. If the organism failed to grow this meant that it had lost its power to synthesize some essential substance. Now, if the mutant will grow on the minimal medium after a certain vitamin or amino acid has been added, this is the substance that the mutant is unable to synthesize. Once a mutant has been discovered it is a simple matter to test it with the various vitamins and amino acids to determine which one is required and hence determine what the normal gene is somehow responsible for synthesizing.

Neurospora is particularly well adapted for this work, not only because it will grow on a chemically defined medium, but also because of the simplicity of its life cycle. The growth requirements are all studied in the haploid gametophyte stage so there is no dominance or recessiveness to complicate the analysis. Also, the four cells that form from the two meiotic divisions are enclosed in a bag, the ascus, where, after a haploid mitotic division, they form eight spores. These ascospores can be tested directly for their genetic properties. This makes it possible for the investigator to observe segregation directly in a single ascus, whereas in an ordinary animal or plant he can only infer it from statistical analysis of Mendelian ratios.

In some cases several genes are found, each of which is required for the synthesis of a certain compound. This enables one to discern the individual steps in the sequence of the synthesis. For example, seven genes were found to be concerned in the synthesis of arginine. Strains lacking four of these genes would grow on ornithine, citrulline, or arginine. Two of the mutant strains would grow on citrulline or arginine, and one would grow only if arginine was added to the medium. Hence the order of synthesis is as follows:

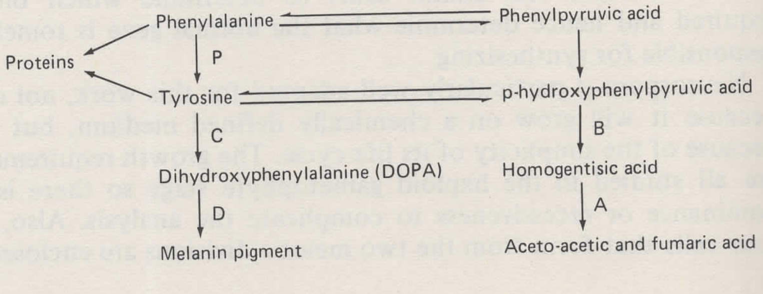

gene 1
gene 2
gene 3
gene 4
NH₂ CH₂ CH₂ CH₂ CH-COOH NH₂
Ornithine
gene 5
gene 6
NH₂ C=O NH CH₂ CH₂ CH₂ CH-COOH NH₂
Citrulline
gene 7
NH₂ C=NH NH CH₂ CH₂ CH₂ CH-COOH NH₂
Arginine

The same synthesis occurs in the human liver. Arginine splits to form a molecule of urea and a molecule of ornithine, which in turn is used to form more citrulline and arginine. In this way a small amount of ornithine is used over and over again to produce urea.

Some of the same mutants are known in the human. For example, there is a recessive disease in which the step from citrulline to arginine is partially blocked. This must be analogous to a mutation of gene 7 in the diagram.

PHENYLALANINE METABOLISM

An example of a series of chemical processes that take place in the human body is given below. Phenylalanine and tyrosine are common amino acids that are part of the normal diet and are normal constituents of protein. They are converted along the paths in the diagram by enzymes, some of which have been shown to be under the control of specific genes.


Phenylalanine
Phenylpyruvic acid
Proteins
P
Tyrosine
p-hydroxyphenylpyruvic acid
B
C
Dihydroxyphenylalanine (DOPA)
Homogentisic acid
A
D
Melanin pigment
Aceto-acetic and fumaric acid

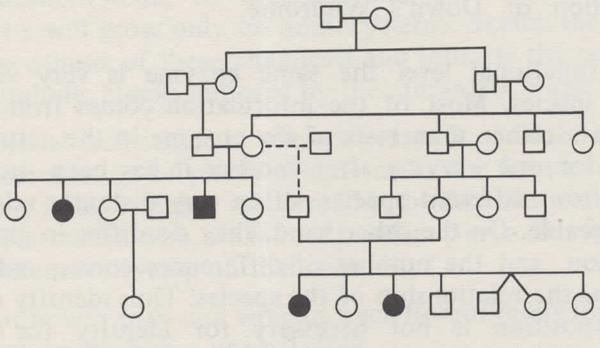
There may be other paths by which these compounds are converted into one another, but these are thought to be the main ones. Also, what is indicated here as a single step may involve several intermediate compounds.

I have already mentioned reactions C and D which are part of the pathway leading from tyrosine to melanin. Reaction C is catalyzed by the enzyme tyrosinase, which is produced by the dominant gene *C*. The genotype *cc* results in the failure of this reaction and the person is albino. Albinism can also be produced by the failure of step D, as was noted for rodents in the previous chapter.

Some persons lack the enzyme necessary to carry out reaction A. When this happens, homogentisic acid accumulates instead of being converted into further compounds. It is then excreted in the urine, where it is easily detected because it forms a dark pigmented substance on exposure to air. Since homogentisic acid is also called alkapton, the condition is known as alkaptonuria. It is inherited as a recessive trait, and was one of the first of Garrod's inborn errors of metabolism. The dominant normal gene *A*, produces the necessary enzyme.

When reaction B is missing, tyrosine and p-hydroxyphenylpyruvic acid are excreted. This is also a rare metabolic disease.

The failure of step P leads to the recessive disease phenylketonuria (PKU) characterized by the accumulation of phenylalanine in the blood and the excretion of this and phenylpyruvic acid. Persons with PKU are severely mentally retarded, so somehow the normal metabolism of phenylalanine is necessary for mental development. The pedigree below shows clearly the recessive inheritance of PKU.



Note the familial occurrence, and the consanguinity among the parents of affected persons.

The successful treatment of PKU is one of the early triumphs of biochemical genetics. If the disease is detected soon after birth the child can be put on a phenylalanine-free diet, thus preventing mental retardation. Phenylpyruvic acid in the urine can be detected by a simple color test, so it is possible to screen large numbers of newborn infants at a small cost.

OTHER INHERITED METABOLIC DISEASES

Now that the pattern is established, new metabolic diseases are being discovered at a rapid rate. These are typically rare and recessive, the normal allele being responsible for the necessary enzyme. The number that are well understood now is about 100. Here are a few examples.

A baby with galactosemia cannot tolerate milk, because it lacks an enzyme (transferase) that converts galactose-1-phosphate into glucose-1-phosphate. The child gets along fine if fed a milk substitute that does not contain lactose. The heterozygotes have half the normal level of enzyme, which is sufficient for normal galactose metabolism, but which permits a chemical analysis based on enzyme level. Thus when two parents are discovered to be heterozygous, the physician can be on the lookout for an affected child and start treatment promptly.

Tay-Sachs disease is also recessive. In this case the deficient enzyme is hexoseaminidase. The disease manifests itself in infancy and the child gradually develops visual, mental and physical impairments that cause death in the first few years of life. For some reason the gene is much more common in the Jewish population than elsewhere. The carrier state can be detected by having only half the normal enzyme level. Furthermore, the enzyme can be studied in the amniotic fluid, so prenatal detection of homozygous embryos is possible. Successful programs to detect families where both parents are heterozygous and then do embryonic testing in such families have been started in several cities.

Such tests on amniotic cells usually require more cells than can be gotten by amniocentesis without danger to mother or fetus. But the cells can be grown in culture and the test made after they have multiplied. By 1973 more than 30 metabolic diseases could be detected in the embryo by this method. This procedure is also very useful for the study of chromosomes, as in the prenatal identification of Down's syndrome.

THE SIMILARITY OF CORRESPONDING ENZYMES IN DIFFERENT SPECIES

On a functional level the same enzyme is very similar in different species. Most of the information comes from *in vitro* experiments rather than tests of the enzyme in the actual organism, but for one enzyme after another it has been shown that enzymes from different species—often very distantly related—are interchangeable. On the other hand, they do differ in amino acid composition, and the number of differences corresponds to the distance of the relationship of the species. Thus identity of amino acid composition is not necessary for identity (or at least extremely close similarity) of function.

ISOZYMES

It is likely that an organism is capable of producing a thousand or more different kinds of enzyme proteins. The number of metabolites in *E. coli* involved in known chemical processes is about 600-800. The number of different proteins could be 2000 or more. In higher organisms the number may well be higher, but there is no reason to think that it is enormously greater.

seases are
rare and
necessary
bout 100.

se it lacks
hate into
l a milk
otes have
r normal
ysis based
ed to be
affected

deficient
n infancy
al impair-
For some
opulation
ving only
e can be
nozygous
ies where
testing in

than can
or fetus.
after they
could be
re is also
prenatal

similar in
in *in vitro*
ual organ-
own that
lated—are
mino acid
ds to the
of amino
at least

thousand
umber of
ocesses is
be 2000
igher, but

The technique of gel-electrophoresis has made it possible to distinguish between proteins that are identical in function and in gross structure by their differential speed of migration in an electric field. This has led to the discovery of **isozymes**, enzymes that are functionally indistinguishable but which have minor chemical differences due to amino acid replacements at a small number of sites. This shows clearly that enzyme function does not depend on an exact amino acid sequence. Rather, there are many sites where the amino acids can be changed without altering the function to any noticeable extent.

A systematic search for isozyme variability in natural populations has revealed that such variability is enormous. The typical Drosophila is heterozygous for at least 10% of the loci governing known isozymes, and the same is true for humans. About one-third of the loci are polymorphic. Furthermore, these are minimum numbers since they depend on a single technique.

The "wild-type allele" is very often, perhaps typically, not a single entity but a collection of indistinguishable isoalleles. Whether these are exactly equivalent in function, so that even natural selection over long geological periods does not prefer one to another, or whether they are only so similar as to not be distinguishable by our relatively crude measurements, is not known.

PROBLEMS

1. A new mutant in Neurospora requires chemical X for growth and accumulates chemical Y. What can you say about the order of these substances in the biosynthetic sequence, and where in this chain does this gene act?
2. In Neurospora, mutant *x* will grow if provided with cystathionine, homocysteine, or cysteine; mutant *w* will grow if provided with homocysteine or cystathionine, but not on cysteine. Mutant *r* will grow only on homocysteine. Sketch the sequence of biosynthesis of these chamicals and indicate the positions of the metabolic blocks imposed by the mutants *x*, *r* and *w*.
3. The mental deficiency in phenylketonuria could be caused by the accumulation of phenylalanine or by the failure to produce tyrosine. What evidence is offered by the fact that children with phenylketonuria who are fed from infancy on a diet very low in phenylalanine develop normal mentality?
4. Explain how a man and woman, each homozygous for recessive albinism, might have a normal child.
5. In Neurospora, which of the following is not expected: (a) a mutant that will grow when ornithine or citrulline is added, but not when arginine is; or (b) a mutant that will grow when citrulline or arginine is added, but not when ornithine is?
6. There are two human mutant phenotypes that show an accumulation of citrulline. Both are caused by homozygous recessive genes. The first type accumulates an even larger amount of

arginosuccinic acid, but the second accumulates only citrulline. Neither accumulates arginine. Where is arginosuccinic acid in the ornithine cycle and at what steps are the mutant genes located?

7. All the known human mutants in the ornithine cycle cause only a partial inactivation rather than a complete inactivation of the enzymes. Suggest an explanation.

8. The enzyme converting ornithine to citrulline is produced by an X-linked gene. Affected males die in early infancy. Heterozygous females are highly variable, some being essentially normal while others are severely affected. Suggest an explanation.

DNA, RNA, AND PROTEIN

12

REPLICATION, TRANSCRIPTION, AND TRANSLATION

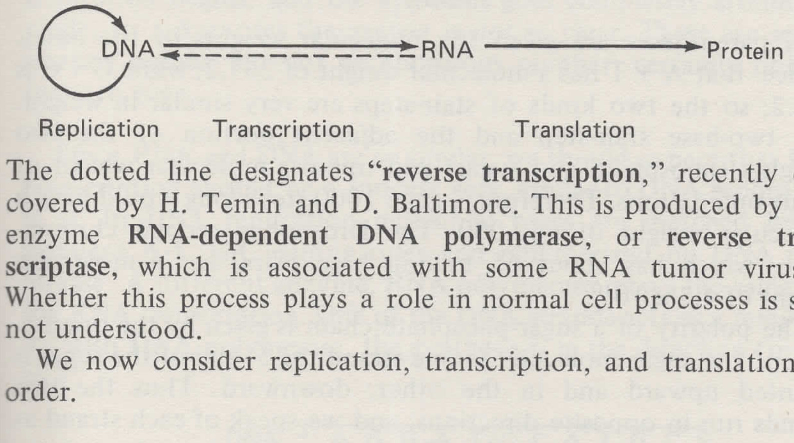
The last chapter told us that enzymes are under the control of genes. Enzymes and other proteins are the key to much that happens in biology, so a central problem for genetics is how the DNA influences protein synthesis.

There is one immediate difficulty: DNA is in the nucleus of the cell, while protein synthesis takes place in the cytoplasm. There has to be some way for the genetic message of the DNA to be carried to the cytoplasm. This is done by another nucleic acid, **ribonucleic acid (RNA)**, which serves as the messenger from the DNA to the protein-synthesizing machinery. RNA is transcribed from DNA in much the same way as DNA copies itself. It is very similar to DNA and can pair with a single strand of DNA to form a hybrid molecule. RNA differs from DNA in two ways:

1. Thymine (T) is replaced by **Uracil (U)**, another pyrimidine, and
2. Deoxyribose is replaced by ribose.

In some viruses, RNA is the genetic material instead of DNA.

The basic principles are encompassed in the following diagram, which illustrates the "Central Dogma."

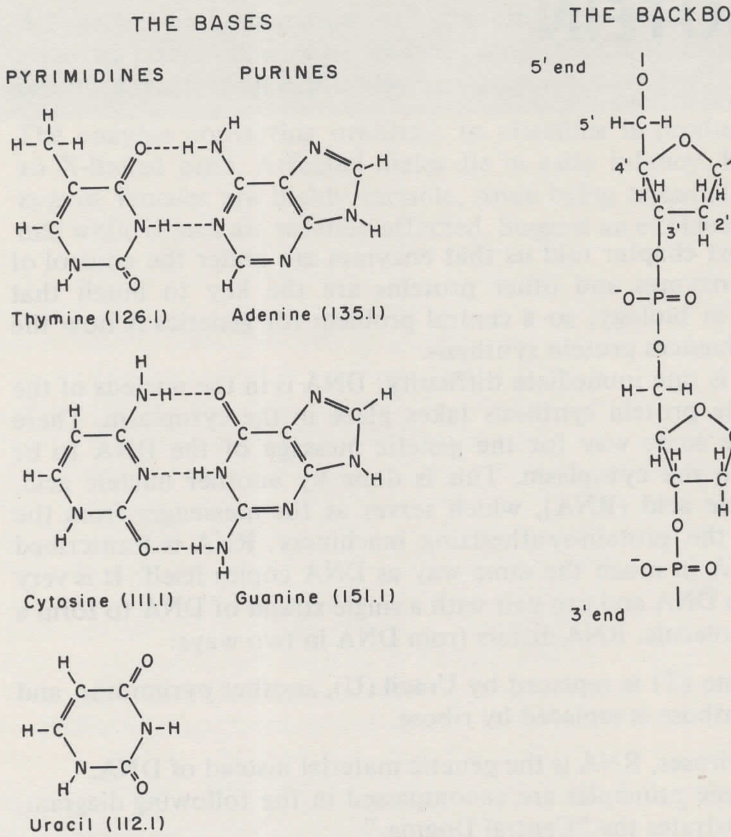


The dotted line designates **"reverse transcription,"** recently discovered by H. Temin and D. Baltimore. This is produced by the enzyme **RNA-dependent DNA polymerase,** or **reverse transcriptase**, which is associated with some RNA tumor viruses. Whether this process plays a role in normal cell processes is still not understood.

We now consider replication, transcription, and translation in order.

A MORE DETAILED LOOK AT THE DNA MOLECULE

The four bases of DNA (along with uracil for RNA) are shown on next page along with the sugar-phosphate backbone. Notice, as mentioned in Chapter 2, that A and T form two hydrogen bonds (indicated by dotted lines) while G and C form three. Note also that U is very similar to T, and it is therefore natural that U can replace T in RNA. A purine or pyrimidine together with the 5-carbon sugar is called a **nucleoside**. When the phosphate is included in addition, it is called a **nucleotide**.



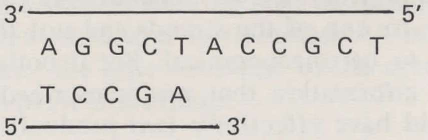
In parentheses are given the molecular weights of the bases. Notice that A + T has a molecular weight of 261.2, while G + C is 262.2; so the two kinds of stair-steps are very similar in weight. The two-base stair-step and the adjacent portion of the two uprights (phosphate and sugar) add up to a molecular weight of about 660 daltons. Therefore, a gene 1000 steps long would have a molecular weight of 660,000. The Drosophila second chromosome, which has about 65 million base pairs, has a molecular weight of 4.3×10^{10}.

The polarity of a sugar-phosphate chain is given by the orientation of the sugar molecule. In one strand the 5′ end of the sugar is oriented upward and in the other, downward. Thus the two strands run in opposite directions, and we speak of each strand as having a 5′ and a 3′ end. (The reason for the primes is to distinguish the atoms in the sugar from those in the bases.)

The two strands are twisted into a helix which makes a complete 360° turn every 10 steps. The distance from step to step is 3.4×10^{-7} mm, so the DNA of the Drosophila second chromosome is about 22 mm long.

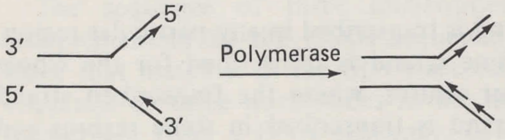
DNA REPLICATION

There is an enzyme, **DNA polymerase**, found by Arthur Kornberg, that can synthesize a strand of DNA complementary to an existing strand. This builds the new strand in the direction 5′ to 3′.



It would be natural, if the model suggested in Chapter 2 were correct, for there to be another enzyme that synthesizes in the opposite direction. Then both strands of the DNA could be replicated at once. But despite a great deal of research in this area, such an enzyme has not been found, so the question of DNA replication is still unsettled.

One possibility is that another enzyme, **ligase**, is involved. This enzyme will join pieces of DNA together. So the process could go like this:

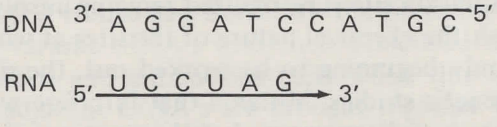


Evidence for this, or some similar process, comes from the fact that during DNA synthesis short segments of newly synthesized DNA can be recovered. But I should emphasize that this is only one of several suggested models, and the true mechanism is not known.

In the bacterial chromosome there is a particular point at which replication begins, and the synthesis goes completely around the circle. In eukaryotes the answer is not so clear. There are several starting points, but not an enormous number; certainly not one for each gene.

TRANSCRIPTION

Since RNA and DNA are so similar, we should expect that RNA transcription should be a process very similar to DNA replication. It is. In DNA replication individual bases are matched by the standard A-T, G-C pairing rules and polymerized by DNA polymerase. A different enzyme, **RNA polymerase,** plays the same role for RNA transcription. One of the DNA strands acts as a template. As with DNA polymerase, the synthesis is in the direction from 5′ to 3′.



The RNA that is made by this transcription process and which carries the message to the protein synthesizing machinery is called **messenger RNA**, or **mRNA**.

Only one of the two DNA strands serves as a template. This has been demonstrated in several ways. One of the most direct was to utilize a DNA that had many more purines in one strand and pyrimidines in the other; the RNA could then easily be seen to be complementary to one of the strands and not the other. At first this may seem to be uneconomical. But if both strands of DNA carried genetic information that was transcribed into RNA, then each gene would have effectively two products with very likely quite different functions. Simultaneous evolution of both would be very difficult to coordinate. So, we are not surprised that only one strand is informational. We shall refer to the strand that serves as the template for transcription as the "sense" strand. The other we shall call the "anti-sense." Thus if the

anti-sense strand is 5′ A T G T T G C A T 3′, the
sense strand is 3′ T A C A A C G T A 5′, and the
transcribed RNA is 5′ A U G U U G C A U 3′.

Although only one strand is transcribed in any particular region, it is not true that the same strand is transcribed for the whole chromosome. In the larger viruses, where the transcribed strand can be identified, one strand is transcribed in some regions and the other strand in other regions. Presumably this is also true of higher organisms. The correct statement is that only one strand is transcribed in a given chromosome region. Of course, if the strand being transcribed switches in different parts of the chromosome, then the direction of transcription must reverse, for it always proceeds so the RNA is synthesized in the 5′ to 3′ direction.

The process of transcription can start at very many places along the chromosome. A distinct mRNA is made for each gene or small group of genes. This raises the question of how the transcribing enzyme can recognize the proper starting point. This appears difficult because of the uniform double-helical structure of DNA; there are no landmarks for the enzyme to recognize. One suggestion comes from analogy to restriction enzymes that cut DNA at very specific places. As mentioned in Chapter 9 (see especially problem 9), these are known in many cases to recognize a reverse repeat in the DNA sequence. For a while it was suspected that this might be the way in which the site for initiation of transcription was recognized, but more recent studies have failed to find any such DNA sequences. So the problem of how the initiation site is recognized remains unsolved.

Although the chemical nature of the sites at which transcription begins is only beginning to be worked out, the site has been identified by genetic studies. Mutants that interfere with the transcription process have been mapped at the position where transcription begins. This region is called the promoter and will be discussed further in Chapter 18.

The transcription process is quite rapid, being some 50 nucleotides per second in *E. coli*. This is much slower than DNA replication under conditions of rapid growth.

The mRNA molecule is long (except in comparison to DNA), usually with a few thousand nucleotides. As stated before, its function is to carry the gene's message to the cytoplasm where it can participate in protein synthesis.

TRANSFER RNA

Transfer RNA, or **tRNA**, is transcribed from the DNA in the same way as mRNA. Transfer RNA molecules are much smaller, having 70 or 80 nucleotides. Like mRNA the molecule is single stranded. However, it carries a number of complementary sequences so that when these pair up the molecule is shaped something like a clover leaf.

Each tRNA has the capacity to attach to a specific amino acid. This is done at the "stem" of the clover leaf. At the end of the leaf opposite the stem is a group of three nucleotides. These three can match up with the complementary three in the mRNA.

The sequence of three nucleotides in the mRNA that is complementary to the tRNA sequence is called the **codon**. The three that match it in the tRNA are called the **anticodon**. Thus a sequence of three nucleotides in the mRNA, which in turn were transcribed from a DNA sequence, determine a particular amino acid.

TRANSLATION

We should first realize that the three-dimensional shape, and therefore the function, of a protein is determined solely by the sequence of amino acids. The sequence of amino acids in the polypeptide thread determine, through secondary attractions among the various kinds of amino acids, the way in which the thread twists and bends to form a complicated three-dimensional structure. This came as something of a surprise to many biologists and chemists, who thought that the linear sequence was not sufficient to determine the complex shape without the aid of some other structure to serve as a template. A definitive experiment was performed by C. Anfinsen, who was able to take a functional protein with a characteristic three-dimensional structure, unravel it into a single thread, and then let it spontaneously reassociate. When it did this, it not only had the physical properties of the original protein, but also the biological activity. So the problem of protein synthesis is the translation of a linear DNA message into a linear amino acid sequence.

As the mRNA moves into the cytoplasm, it becomes associated with particles called **ribosomes**. Somehow the ribosome guides the mRNA and the various tRNA's with their associated amino acids so that the codons and anticodons pair. When this is done the amino acids of adjacent tRNA's are brought together where they can join together to make a polypeptide, as if the ribosome were a jig.

The mRNA moves through the ribosome much as a tape recording moves through the playing heads. Often several ribosomes attach to the same mRNA so that several identical polypeptides are synthesized simultaneously. The ribosome first attaches at the 5′ end and moves toward the 3′ end. In eukaryotes the direction is not so important, but in bacteria where there is no clear distinction between nucleus and cytoplasm, this means that the front end of a messenger RNA can begin protein synthesis while the tail end of the same molecule is still being transcribed.

The amino acids are joined together in the polypeptide chain so that the amino end of the molecule is completed first and the carboxyl end last. This was first demonstrated experimentally by using radioactively labeled amino acids in a protein-synthesizing system. The amino acids that were added latest were at the COOH end of the polypeptide.

We can summarize this process as follows:

$$\begin{array}{lll}
\text{DNA} & \left\{\begin{array}{l}\text{anti-sense}\\ \text{sense}\end{array}\right. & \begin{array}{l}5'\ \text{A}\ \text{T}\ \text{G}\ \text{T}\ \text{T}\ \text{G}\ 3'\\ 3'\ \text{T}\ \text{A}\ \text{C}\ \text{A}\ \text{A}\ \text{C}\ 5'\end{array}\\
\text{mRNA (two codons)} & & 5'\ \text{A}\ \text{U}\ \text{G}\ \text{U}\ \text{U}\ \text{G}\ 3'\\
\text{tRNA (two anticodons)} & & 3'\ \underbrace{\text{U}\ \text{A}\ \text{C}}\ \underbrace{\text{A}\ \text{A}\ \text{C}}\ 5'\\
\text{Amino acids} & & \text{NH}_3\quad \text{Met}\quad \text{Leu}\quad \text{COOH}
\end{array}$$

The amino acid methionine was attached to a tRNA with the anticodon 3′ UAC 5′, which corresponded to the codon 5′ AUG 3′ on the messenger RNA molecule, which in turn was produced from a DNA sense strand with the sequence 3′ TAC 5′. Likewise leucine was ultimately determined by the DNA sequence AAC.

This illustrates the genetic code for two amino acids. In the next chapter we shall see what the code is for all the amino acids.

THE RIBOSOME AND RIBOSOMAL RNA

The ribosome consists of two parts, one large and one small. At the beginning of protein synthesis these two join, and somehow their three-dimensional structure aids in bringing together the mRNA and appropriate tRNA's and in making the peptide linkages between the amino acids that have been brought together. After the synthesis is complete, the two parts come apart, to rejoin with another mRNA and start the process again.

It is no surprise that ribosomes are made of protein and nucleic acid. In this case the nucleic acid is a third kind of RNA, **ribosomal RNA** or **rRNA**. This is transcribed from genes just as other RNA's are. It is interesting that a classical mutant of Drosophila, called "bobbed" because of the reduction of some of the bristle lengths, is caused by a deficiency of rRNA. The mutant causes many abnormalities including slow development, not surprising in view of the fundamental defect.

The ribosome also contains about 50 kinds of proteins which are synthesized in the same way as other proteins. How are these proteins and RNA put together to make the complex ribosomal structure? The answer is very striking: they put themselves together! M. Nomura was able to do the critical experiment of taking ribosomes apart into their individual proteins and RNA's. If left alone, these units spontaneously fit together and produce a functional ribosome. So all the necessary information is contained in the rRNA and the amino acid sequences of the 50 proteins.

Now that ribosomes can be studied in detail, we should soon know how they carry out their remarkable assignment. I will give one example of the kind of work that has been done. It is known that streptomycin interferes with protein synthesis by causing errors in translation. Some strains of bacteria are resistant to streptomycin, and M. Nomura was able to show that this is caused by a change in a particular one of the proteins. This research has helped us to understand how the drug works and how bacteria became resistant to it.

SUPPRESSOR GENES

Suppressor genes were mentioned briefly in Chapter 10 and an example from Drosophila was given in problem four of that chapter. This was a suppressor of a specific mutant gene, the suppressor being located at another locus. That a second mutant should sometimes undo the effect of the first is not particularly surprising, and it might have a number of possible chemical explanations, such as opening up an alternative pathway for synthesizing the normal gene product.

More puzzling at first is an intragenic suppressor, a second mutant in the same gene that restores the gene function but is not simply a reversal of the original DNA change. In higher organisms these are frequently indistinguishable from back mutations, but in microorganisms they can be shown to involve a different nucleotide by recombination tests. One possibility is that the original mutation is a duplication and the suppressor is a deletion of the same length, usually a single nucleotide; or vice versa. A deletion or duplication can throw the whole gene out of order because all the coding units beyond this one are wrong. For example, if the original gene segment is CAT GAG ACT TAG TAT GAT, a deletion of the fourth nucleotide produces CAT AGA CTT AGT ATG AT which is wrong to the right of the deletion for all the rest of the gene, and this may be hundreds of nucleotides. However, a later duplication can throw it back into order. For example, a duplication of T can give CAT AGA CTT TAG TAT GAT, where the nucleotide added by duplication is underlined. Only two amino acids are wrong in the whole polypeptide, so it may function normally or approximately so.

A second explanation is that the first mutant somehow changes the three-dimensional structure of the protein so that it is inactive or otherwise abnormal. A second amino acid change at another position may just happen to warp the protein molecule so that it resembles the original shape and therefore functions approximately normally.

Still more puzzling to geneticists have been generalized suppressors—genes that suppress a number of seemingly unrelated mutants. A surprising feature is that certain mutants at a number of loci are suppressed but other mutants at the same loci are not. An explanation for at least some of these is to be found in the translation process. These suppressor genes cause mistakes in the amino acid incorporated into the polypeptide. Usually there is a specific misreading so that the original three letter code is "read" part of the time as if it were something else. Under some circumstances two wrongs make a right by having the misreading of the wrong sequence lead to the same amino acid as the correct reading of the right sequence. Many suppressors in *E. coli* have been shown to work this way. It might appear that there would be so many mistakes made from normal sequences as to more than offset any benefit from repairing the mutant, but at least in some cases a key enzyme may be needed in very small amounts so that only a small proportion of mistakes is enough to produce a few needed molecules. However, we are not surprised that mutants suppressed in this way are usually not fully normal organisms.

A BIT OF HISTORY

The first identification of a specific protein defect with a mutant gene was for the hemoglobin of sickle cell anemia. Linus Pauling showed in 1949 that hemoglobin S differed from normal hemoglobin A in its electrophoretic mobility. Later, in 1957, this difference was shown by Vernon Ingram to be caused by the substitution of neutral valine for the negatively charged glutamic acid. The relevant part of the 146 amino acid molecule is shown below, along with that for another abnormal hemoglobin. The rest of the molecule is identical in all three.

Table 12-1

Amino acid number	Hemoglobin A	Hemoglobin S	Hemoglobin C
1	valine	valine	valine
2	histidine	histidine	histidine
3	leucine	leucine	leucine
4	threonine	threonine	threonine
5	proline	proline	proline
6	GLUTAMIC ACID	VALINE	LYSINE
7	glutamic acid	glutamic acid	glutamic acid
8	lysine	lysine	lysine

One of the first convincing demonstrations that the gene and its protein product are co-linear was done with *E. coli* by Charles Yanofsky. He identified the amino acid replacement in the protein produced by a series of mutants in the same gene and also mapped the positions of the mutants by fine-structure crossover tests. Some of the results are shown below:

As can be seen, not only are the mutant sites in the right order, but the distances are roughly proportional to the number of amino acids separating them. This work, in the early 1960's, clearly

t a number
oci are not.
ound in the
akes in the
y there is a
le is "read"
nder some
misreading
the correct
. *coli* have
e would be
more than
ast in some
unts so that
duce a few
at mutants
nisms.

ect with a
emia. Linus
om normal
1957, this
sed by the
ed glutamic
le is shown
in. The rest

globin C
ne
e
ine
e
NE
ic acid

gene and its
by Charles
the protein
lso mapped
sover tests.

right order,
er of amino
0's, clearly

Table 12-2

Normal amino acid	tyr	leu	thr	gly	gly	gly	ser
Mutant amino acid	cys	arg	ileu	arg	val	asp	leu
Amino acid sequence number	11	13	19	47	49	70	71
Map distance		.04	.30	.44	.06	.50	.02

established the correspondence between nucleotide and amino acid sequence.

THE GENETIC CODE

13

SIZE OF THE CODING UNIT

As soon as the linear nature of DNA and of polypeptide chains was understood there began to be speculations as to the nature of the correspondence between nucleotide and amino acid sequences. Since there are 20 amino acids and only four bases there obviously cannot be a 1:1 correspondence; nor can a two-letter code suffice for there are only 16 permutations of four bases two at a time. On the other hand with a three-letter unit there are 64 permutations, more than enough. So we can conclude that the coding unit is at least three.

Overlapping codes were quickly ruled out as the amino acid sequences of many proteins were determined, for such codes place too many restrictions on neighboring amino acids.

Evidence that the code is in fact three came from studies of proflavine-induced mutations. These mutations, produce one-base duplications or deletions. Sometimes a combination of three deletion mutations very close together, or three duplications will produce a normal or near-normal phenotype.

This is easily understood, as explained by Francis Crick. A deletion mutation throws the whole reading system out of register so that all amino acids beyond this point will be wrong. However, three such mutants will bring the system back into register again so that only the amino acids between the first and third mutants will be wrong. If this distance is very short the protein may be nearly normal. These experiments strongly suggest that the coding unit, or **codon**, is three nucleotides in length.

EARLY EXPERIMENTS WITH SYNTHETIC MESSENGER RNA

Although it has not been possible until recently to synthesize RNA with a known sequence, it has been possible to make it with any desired percentage composition. Such synthetic RNA's can be used as messengers in cell-free protein synthesizing systems. All that is needed is a cell extract containing ribosomes, transfer RNA's, ATP and the other ingredients for protein synthesis plus a pool of amino acids. In each experiment a different amino acid is labeled, and one can determine which amino acids are incorporated into the polypeptide by radioactivity counts. In one of the earliest experiments, M. Nirenberg found that an RNA composed only of uracil behaved as a messenger and led to the synthesis of a polypeptide containing only phenylalanine. Thus the codon for phenylalanine must be UUU (if the codon is in fact three units long) and the corresponding DNA is AAA. Since it is customary to give codes in terms of messenger RNA sequences, I shall follow this practice from here on.

Nirenberg and Ochoa then did a number of variations on this theme which led to a great deal of information about the code. For example, poly C produced a polypeptide containing only proline and poly A gave polylysine. A polymer of A and C in random order but with a great excess of A gave lysine, of course, but also threonine, glutamine, and asparagine in the polypeptide molecule. This suggests that these correspond to the three codes, AAC, ACA, and CAA, but there is no way of telling which amino acid goes with which code.

RNA BINDING EXPERIMENTS

In 1964 Nirenberg discovered another trick. He found that, under some circumstances, tRNA will attach specifically to the ribosome–mRNA complex. This attachment is specific; for example, with A A A A A A A messenger only lysine RNA will attach. The happy fact that made the crucial experiments possible is that very short messengers as short as three units will cause a specific tRNA to attach. Since trinucleotides can be made in known order the way was open to determine the sequence of the various codons; now a distinction could be made among AAC, ACA, and CAA. Such experiments were not always clear cut and there were many technical difficulties. Nevertheless, within a few months most of the codons were known with considerable certainty.

PROTEIN SYNTHESIS WITH SYNTHETIC RNA POLYMERS

A different approach was used by H. G. Khorana and his coworkers. They have been able to synthesize DNA's which are repeating triplets with a total length of about 12 (e.g., T T C T T C T T C T T C). They also made repeating doublets of about the same length, such as A G A G A G A G A G A G. Somehow the synthetic RNA made from this DNA is much longer and can be used as a messenger for *in vitro* protein synthesis. This makes possible very specific tests for codes. For example the sequence A G A G A G. . . produced a polypeptide in which there was an alternating sequence of glutamic acid and arginine. This agrees with the codes AGA for arginine and GAG for glutamic acid. This is also strong proof for the triplet code. Furthermore a repeating triplet RNA gave completely monotonous polypeptides with only a single amino acid, as expected.

By 1966 the code was known completely. It has been abundantly confirmed since that time as it has become technically possible to synthesize long DNA or RNA molecules of known sequence. The tour de force came with Khorana's synthesis of a DNA molecule corresponding exactly to the known sequence of one of the genes for transfer RNA in yeast. To be sure, this is one of the smallest genes, but it was a heroic chemical achievement.

The mRNA codons are given below:

5′ 1st	2nd				3rd 3′
	U	C	A	G	
U	Phenylalanine	Serine	Tyrosine	Cysteine	U
	Phenylalanine	Serine	Tyrosine	Cysteine	C
	Leucine	Serine	•	•	A
	Leucine	Serine	•	Tryptophan	G
C	Leucine	Proline	Histidine	Arginine	U
	Leucine	Proline	Histidine	Arginine	C
	Leucine	Proline	Glutamine	Arginine	A
	Leucine	Proline	Glutamine	Arginine	G
A	Isoleucine	Threonine	Asparagine	Serine	U
	Isoleucine	Threonine	Asparagine	Serine	C
	Isoleucine	Threonine	Lysine	Arginine	A
	Methionine	Threonine	Lysine	Arginine	G
G	Valine	Alanine	Aspartic Acid	Glycine	U
	Valine	Alanine	Aspartic Acid	Glycine	C
	Valine	Alanine	Glutamic Acid	Glycine	A
	Valine	Alanine	Glutamic Acid	Glycine	G

SOME PROPERTIES OF THE CODE

It is clear from the chart that there is a great deal of **degeneracy**, or better, **synonymy**, in the code. There are several codes for most of the amino acids. It is also clear that the third nucleotide of the codon is less crucial than the first two; in most cases the two purines are interchangeable, as are the two pyrimidines. For example, UAU and UAC both code for tyrosine. In some cases, all four are interchangeable, e.g., threonine.

As the code became understood, a crucial question still had to be answered: is the code different for different species, or is it the same for all organisms? Abundant evidence now proves that the code is universal. For example, the mutational changes in human hemoglobins are consistent with single nucleotide changes within the code as worked out from bacterial systems.

Also, a test tube system with mRNA from a rabbit carrying the message for hemoglobin and everything else from bacteria makes rabbit hemoglobin, so the bacterial system must translate the message the same way the rabbit system does. Another striking line of evidence is that viruses that normally grow in human or mouse cells can be made to grow in bacteria, despite the fact that in each case they use the ribosomes and transfer RNA's of the host. The most striking implication of this is that the code must be very, very old, having started with the ancestors of all the organisms now living.

STOPPING AND STARTING

Three of the codons are indicated by a period. These do not code for a specific amino acid, but are a signal that the message is completed. The synthesis stops and the polypeptide falls off the ribosome.

If a mutation occurs changing one of the normal amino acid codons to a terminator, for example UAC to UAA, then the synthesis will stop at this point and a short polypeptide will be produced. This is one way in which such termination signals were discovered. Conversely, if a terminator mutates to an amino acid codon, the synthesis will run on past the normal length and produce an extra-long polypeptide. This has been one way of finding what DNA beyond the normal protein coding region does.

The starting signal has not been so easy to discover. All the three-letter permutations are used up, so there is no specific sequence to be used as a start signal. Some insight was provided by the realization that the first amino acid in the polypeptide is always methionine (although this is often removed later). There are two distinct tRNA's for methionine, and they differ slightly in their structure. It is also known that one of these binds to the small ribosome component; these then join to the large component to make the full-size ribosome. Somehow, one type of methionine tRNA will only be incorporated at the initiation of translation whereas the other type only fits the full-sized ribosome and therefore can insert methionine only interstitially. This has been shown in specific experiments where the initiator tRNA binds to the small ribosome component and the interstitial tRNA binds to the full ribosome.

There is another difference. The initial amino acid is formylmethionine rather than methionine. This makes sense, because a hydrogen in the amino group is replaced by a formyl group. This blocks the formation of a polypeptide chain, so this can only add additional amino acids in one direction. Thus the same codon, AUG, serves as the code for formyl-methionine when it is at the beginning and for methionine when it is in the middle of the message.

THE WOBBLE HYPOTHESIS

One of the striking things about tRNA's is that they contain some unusual bases, and (even more surprising) that some of these are in the anticodon region. Francis Crick was the first to notice, from detailed consideration of the molecular structures, that the third position in the codon required a less exact fit. He suggested that at this position a slight "wobble" is permitted by the structural constraints. The rules of pairing that he proposed have stood up completely with subsequent work.

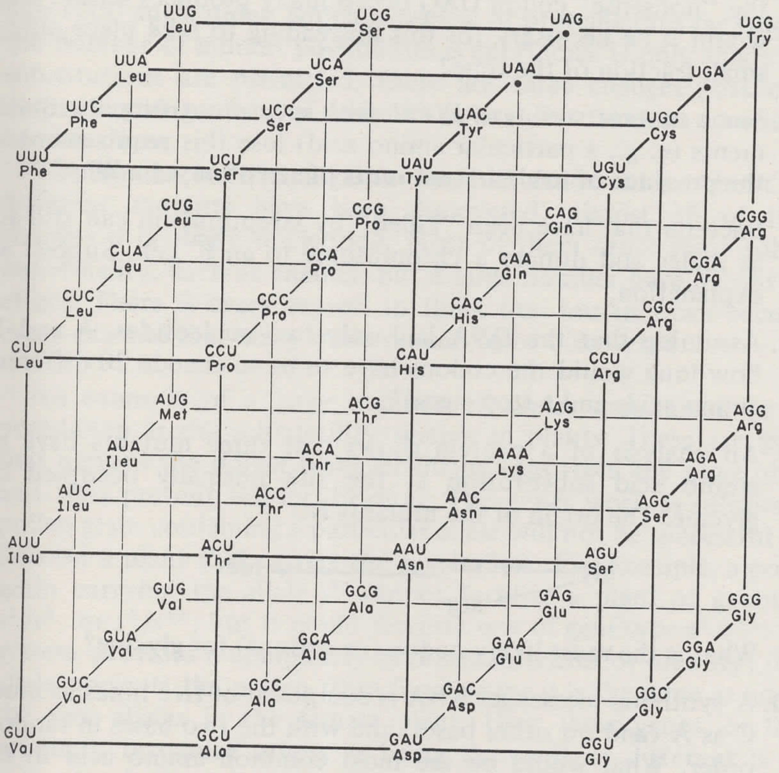
The wobble rules for the third (3′) end of the codon and the 5′ end of the anticodon, in distinction to the normal Watson-Crick rules which apply to the first two positions, are:

Codon	Anticodon
U, C	G
G	C
U	A
A, G	U
A, U, C	I

"I" stands for inosinic acid, a nucleotide we have not yet encountered.

ANOTHER DIAGRAM FOR THE CODE

Here is another way to present the mRNA codons. It has some advantages for seeing the effects of mutations. Any substitution of the (left) letter of the triplet codon causes a change vertically, with the pyrimidines in the upper half and the purines in the lower half. Likewise, a change in the second position causes a change to the right or left, and again a change from a purine to a pyrimidine means crossing the center line. Finally, a change in the third position causes a shift in the direction perpendicular to the page. A change that requires a move other than one of the three main directions cannot be accomplished by a single base change.


UUG Leu
UCG Ser
UAG
UGG Try
UUA Leu
UCA Ser
UAA
UGA
UUC Phe
UCC Ser
UAC Tyr
UGC Cys
UUU Phe
UCU Ser
UAU Tyr
UGU Cys
CUG Leu
CCG Pro
CAG Gln
CGG Arg
CUA Leu
CCA Pro
CAA Gln
CGA Arg
CUC Leu
CCC Pro
CAC His
CGC Arg
CUU Leu
CCU Pro
CAU His
CGU Arg
AUG Met
ACG Thr
AAG Lys
AGG Arg
AUA Ileu
ACA Thr
AAA Lys
AGA Arg
AUC Ileu
ACC Thr
AAC Asn
AGC Ser
AUU Ileu
ACU Thr
AAU Asn
AGU Ser
GUG Val
GCG Ala
GAG Glu
GGG Gly
GUA Val
GCA Ala
GAA Glu
GGA Gly
GUC Val
GCC Ala
GAC Asp
GGC Gly
GUU Val
GCU Ala
GAU Asp
GGU Gly

PROBLEMS

1. Streisinger reports that for a short sequence of the amino acids in the lysozyme locus in bacteriophage the normal and mutant sequences are:

Position	1	2	3	4	5	6
Normal	Ser	Pro	Ser	Leu	Asn	Ala
Mutant	Val	His	His	Leu	Met	Ala

The mutant has a deleted base at the left end of the first position and a duplicated base at the left end of position 6. See if this is consistent with the code as given in the table in this chapter. (Hint: start with Met)

2. Suggest how "nonsense" mutants (changing a codon into a terminator) could be used to show that polypeptide sequence goes from the amino end to the carboxyl end.

3. What is the minimum number of base changes required to change the amino acid from Thr to His? Thr to Asn? Arg to Gly?

4. Yanofsky reported that in *E. coli* there was a mutant in which glycine was changed to arginine. In another mutant the same glycine changed to glutamic acid. Later a recombination between the two mutants gave the normal glycine again. Is this consistent with the known code and the assumption that each mutant represents a single nucleotide change?

5. One suppressor mutant is known to change the reading so that the "nonsense" codon UAG occasionally produces serine. Why would it be necessary for this misreading to take place only a small fraction of the time?

6. Some mutant strains of *E. coli* with special nutritional requirements (e. g., a particular amino acid) lose this requirement in the presence of sublethal amounts of streptomycin. Why?

7. Bacteria that have been "killed" by streptomycin can still act as males and donate a chromosome to an F^- cell. Suggest an explanation.

8. Assuming that the DNA had only two nucleotides, A and T, how long would the codons have to be to encode 20 different amino acids and a stop signal?

9. An analysis of a protein shows that three mutants have an amino acid substitution at the site normally occupied by glycine. The origin of the mutants is:

$$\begin{tikzcd} \text{gly} \arrow[r] \arrow[dr] & \text{val} \arrow[rr] & & \text{met} \\ & \text{arg} & & \end{tikzcd}$$

What is the most likely codon (or codons) for glycine?

10. A synthetic messenger RNA is composed of five times as much C as A (and no other bases) and with the two bases in random order. What would be the most common amino acid in the protein synthesized from this messenger? Next most common?

11. What amino acids could be obtained by a single base substitution in the methionine codon?

12. What is the minimum number of transfer RNA's for isoleucine? What are the anticodons? What is the maximum number? What are the anticodons?

MULTIPLE ALLELES AND GENETIC FINE STRUCTURE

14

MULTIPLE ALLELIC SERIES

It is apparent that if the coding region of a gene is something like 1000 nucleotides, which is typical of proteins thus far studied, the number of mutant possibilities is enormous. If only single base substitutions are permitted, there are three changes possible at each site giving a total of some 3000—and this takes no account of combinations.

The human beta hemoglobin has 146 amino acids. About 100 different mutants have been discovered, almost all of them changes in a single nucleotide. Most of them cause some sort of abnormality, such as anemia, but a large number have no obvious effect. There is every reason to think that mutants can occur at every site and that many more remain to be discovered.

SELF-STERILITY ALLELES IN PLANTS

An example of a large number of alleles maintained in the population is the self-sterility system in plants. These are genes that prevent the pollen from fertilizing eggs from the same plant, and thus prevent self-fertilization. The genes work as follows: A pollen grain containing a particular allele will not be successful if it falls on a plant that carries the same allele. For example, a pollen grain carrying the allele s^1 cannot fertilize a plant of genotype s^1/s^2, or s^1/s^{12}, but it could fertilize one of genotype s^5/s^{13}. The system provides a completely accurate criterion of identity; if an allele prevents the pollen from functioning it is the same as one of the two alleles in the female plant. How these genes, or their products, recognize others as being the same or different is not understood, but it is clear what their function is. They prevent the harmful effects of self-fertilization. Every breeder knows the weakening effect of inbreeding. This is the plant's mechanism for preventing the most extreme form of inbreeding.

Such a system tends to maintain a maximum number of alleles in the population. If an allele is very common, pollen grains carrying it will frequently fall on flowers carrying the same allele and be nonfunctional. Therefore, these genes would tend to be reduced in frequency by natural selection, since pollen carrying rarer alleles would be more successful in pollination. Hence, rare alleles tend to increase and common alleles to decrease. Finally there will be an equilibrium in which all alleles are equally frequent. Furthermore, each new mutant has an immediate advantage, since a pollen carrying it can fertilize any plant. So the population tends to retain every new mutant that arises. The only counteracting factor is the random loss of rare alleles, especially in small populations.

A large population may have hundreds of different alleles. In one population of about 500 evening primroses, 37 different alleles were found. A large population may have hundreds. Of course, self-sterility alleles are exceptional in that there is strong selection for genetic diversity; but they show the extent of mutational variability that is available.

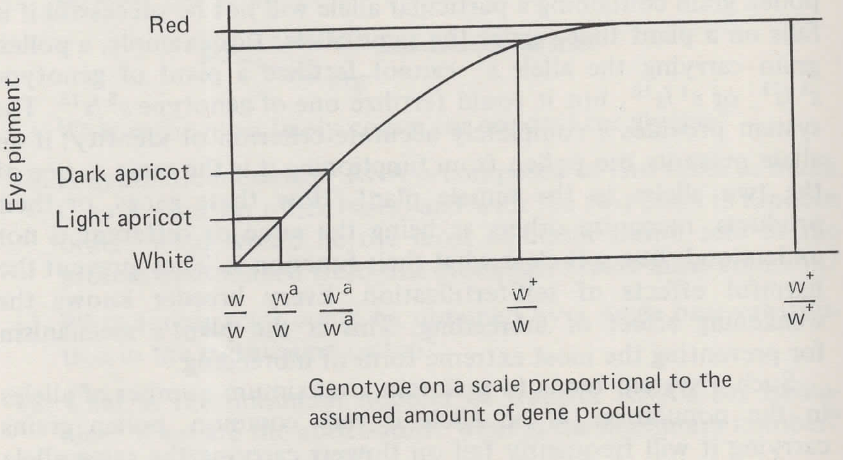
PHENOTYPIC INTERACTIONS OF THE WHITE-EYE ALLELES IN DROSOPHILA

Some 20 or more different color mutants have been identified at the *white* locus in Drosophila. The colors range from nearly normal red to white. They are designated in classical notation by the basic letter identifying the locus, w, with superscripts specifying the particular allele. For example: w^{bl} (blood), w^e (eosin), w^a (apricot), w^h (honey), and w (white).

The white-eye series can again serve as a typical example of dominance relations between pairs of genes in a multiple allelic series. The wild-type gene is dominant (or almost completely so) to any of the other alleles. On the other hand, the intermediate colored alleles show a lack of dominance with each other. In general, the darker the eye color, the more the dominance.

For example, w^+/w, w^+/w^a, w^+/w^a, etc., are all red eyed. But w^a/w is about midway in color between w^a/w^a and w/w.

This relationship makes sense if we assume that the genes are acting to produce an enzyme that catalyzes the eye pigment reaction. A typical curve of diminishing returns is shown below, with a possible interpretation of the action of these genes.



Because of the shape of the curve one w^+ gene produces practically as much eye pigment as two. On the other hand, the apricot gene whose activity is much less lies in the position where two genes produce almost twice as much pigment as one. A gene whose pigment producing activity is intermediate between apricot and w^+ would be of intermediate dominance; two genes would produce more pigment than one, but not twice as much.

MUTANT GENE ACTION IN TERMS OF THE NORMAL ALLELE

The relationship between alleles is usually expressed in terms of dominance. We say that allele A_1 is dominant to A_2 if A_1A_1 and A_1A_2 are indistinguishable phenotypically. If A_1A_2 is half way between A_1A_1 and A_2A_2 there is no dominance; or if A_1A_2 is closer to A_1A_1, A_1 is partially dominant. In some instances A_1A_2 is outside the range of the two homozygotes, in which case there is **overdominance.** With respect to most mutant alleles that have been studied, the wild-type allele is nearly, but not completely dominant. For example, mutants that are lethal when homozygous commonly cause a reduction of viability in heterozygotes of about 5%.

In the previous section we have interpreted the eye color phenotype as if the w^+ and w^a alleles were both producing pigment, the former being more active. On the other hand it could be that the w and w^a alleles are actively reducing the amount of pigment. H. J. Muller has devised techniques to distinguish between these (and other) possibilities and has introduced a suitable vocabulary for describing the phenotypic effect of an allele in terms of that of the wild-type allele.

If a mutant allele produces no detectable effect we say it is an **amorph.** The operational definition of an amorph is that it is equivalent to a deletion; a chromosome with this allele has the same effect as a chromosome in which this region is missing entirely. An example of an amorph is the white-eye allele.

The apricot (w^a) mutant is a **hypomorph**; that is, it acts in the same direction as the normal allele, but less effectively. The evidence for this is that w^a/w^a is more normal (produces more eye pigment) than w^a/w, which is more normal than w/w. Since w is known to be amorphic, the w^a allele is acting in the direction of the normal allele.

An **antimorph** is a mutant with activity in a direction opposite to that of the normal allele. The mutant, abnormal abdomen (A), is such an example:

$$\frac{A}{A} > \frac{A}{\text{deletion}} > \frac{A}{+} > \frac{+}{+}$$

where the symbol $>$ means "is more abnormal than." Somehow the A gene is antagonistic to the normal gene, since the mutant gene causes a phenotype that is farther from normal than that caused by a deletion. Compare the three bottom chromosomes on the left in the example above.

In addition, Muller has suggested two other possibilities: a **hypermorph** (more active than the normal allele) and a **neomorph** (an activity qualitatively different from that of the normal allele).

The great majority of mutants studied have been hypomorphs or amorphs, so the usual result of mutation seems to be a partial or complete inactivation of the gene rather than a change to a new function.

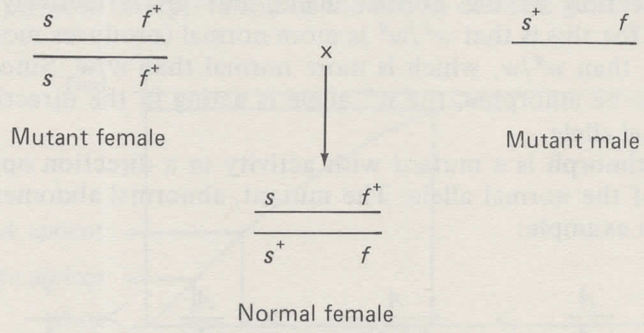
The words hypomorph, amorph, antimorph, hypermorph and neomorph are used to describe relationships between alleles *based entirely on phenotypic manifestations.* Just as with dominance, as discussed in Chapters 10 and 11, these relationships are now beginning to be understood in molecular terms.

Molecular biology has usually dealt with qualitative differences rather than relationships of the kind discussed here, but we can expect eventual understanding of these phenotypic phenomena in molecular terms.

CLASSICAL TESTS FOR ALLELISM

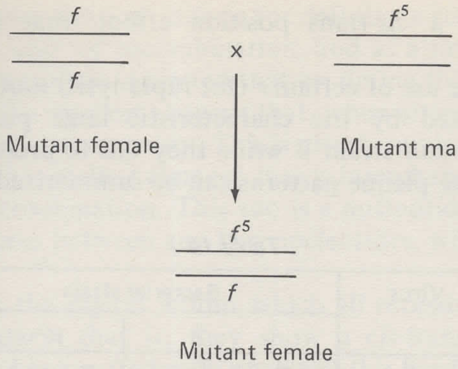
Members of an allelic series almost always are similar in some respect; they all affect the eye pigment, or they all are concerned with wing shape, or they are all concerned with the synthesis of the same chemical. If two mutants occur that affect a similar function, they may be suspected of being alleles.

Usually, though not always, recessive mutants may be identified as allelic or not by a test of complementarity. For example, two mutants on the X chromosome of Drosophilia both cause similar misshapen bristles; they are called forked (*f*) and singed (*s*). If a homozygous singed female is crossed with a forked male, the female progeny have normal bristles. The forked chromosome carries the normal allele of singed; the singed chromosome carries the normal allele of forked. The cross may be diagrammed as follows:



The forked and the singed chromosomes are complementary in the sense that each carries a normal allele of the other. This shows that forked and singed are mutant forms of different genes.

On the other hand, if forked and singed were allelic mutants the female offspring should have been mutant instead of normal, since neither chromosome would have carried the normal allele of the other. This can be illustrated with another mutant that looks very much like singed, but was found by crossing-over tests to be an allele of forked. It is called *forked-5* (f^5). If the same kind of cross is made as before, the results can be diagrammed this way:



Thus, as a rule, if the hybrid between two recessive mutants is abnormal, they are alleles; if the hybrid is normal, the mutants are at different loci.

CIS-TRANS POSITION EFFECT

One of the most striking features of two mutants of the same gene that show recombination is that they show a **position effect**. Two mutants, g^1 and g^2, each causing a garnet eye color are typical. The crossover tests show that g^1 is to the left of g^2. The genotypes and phenotypes are as follows:

$$\frac{+\qquad +}{g^1\qquad g^2}\quad \text{Normal red eye}$$

$$\frac{+\qquad g^2}{g^1\qquad +}\quad \text{Mutant garnet eye}$$

By analogy with chemical terminology, the upper type is called cis and the lower trans. If these differ in phenotype, as in the present case, there is said to be a **cis-trans position effect.** Ordinarily the cis type is normal and the trans type is mutant.

We can make sense out of these results by regarding the whole region as a single gene and letting g^1 and g^2 be two sites where mutational changes have occurred. In the upper diagram the phenotype is normal because there is one intact normal gene, which because of its dominance leads to a normal phenotype. In the lower genotype neither gene is normal, so the phenotype is mutant.

FINE STRUCTURE ANALYSIS IN MICROORGANISMS

Fungi, bacteria, and viruses are particularly useful in studying the fine structure of chromosomes. Large numbers can be grown with ease, and there are selective methods that permit a small number of crossovers to grow and be detected in the midst of millions of noncrossovers. The work of S. Benzer on bacterial viruses is particularly elegant.

Since viruses are haploid, a cis-trans comparison is not possible in the usual way. However, doubly infected bacterial cells accomplish the same purposes. If a bacterial cell with two viruses, *a*+ and +*b*, gives a different phenotype than one with + + and *a b*, this is equivalent to a cis-trans position effect, that is, noncomplementation.

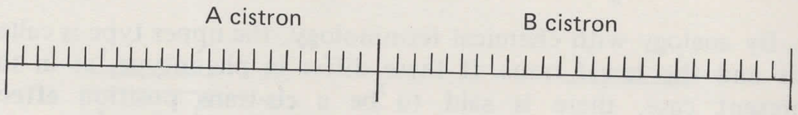
Benzer made use of certain *r* (for rapid lysis) mutants, which are easily recognized by the characteristic large plaque that they produce on *E. coli* strain B while they fail to produce any plaque on strain K. The plaque patterns can be summarized as follows:

Table 14-1

Virus	Bacterial strain	
	B	K
+	Small plaque	Small plaque
r	Large plaque	No growth

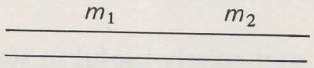
With two different *r* mutants a complementation test consists of doubly infecting bacteria of strain K. If there is complementation there is phage production; if not there is none. A recombination test consists of growing two *r* mutants on B, then testing the progeny on K. If there has been a recombination leading to wild type, there will be growth on K; otherwise no growth. The test is very sensitive and recombination frequencies as low as one in a million or less can be detected.

Benzer analyzed several thousand mutants. They fell very sharply into two complementation groups, A and B, called cistrons A and B, each of which contained hundreds of mappable sites separable by recombination.



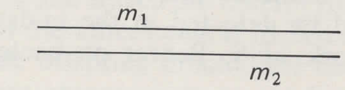
If two mutants were in the A cistron, they did not show complementation. Nor did two mutants in the B cistron. But if one mutant was in the A cistron and the other in the B, they did complement.

The name cistron comes from the fact that within a cistron the cis and trans forms had a different phenotype. For example,



has a normal phenotype

whereas



has the mutant phenotype

We can understand this easily if we realize that a cistron is the region that codes for one polypeptide. In the upper diagram, there is one normal polypeptide produced, but in the lower diagram there is none.

There are three classical criteria for defining a gene: as a unit of mutation, as a unit of recombination, and as a unit of function. Accordingly Benzer has suggested that we define three entities:

A **Muton** is the smallest element that, when altered, gives rise to a mutant form of the organism. This is clearly a nucleotide.

A **Recon** is the smallest element that is interchangeable, but not divisible, by recombination. This too is a nucleotide, since recombination can occur between any two nucleotides, within or outside a gene.

A **Cistron** is the region within which all recessive mutants are noncomplementary; that is, they show a cis-trans effect. In all cases that have been sufficiently analyzed this has been the region that codes for a polypeptide.

The experiments reported in this chapter were all done by purely genetic methods, with little aid from cytology or chemistry. They illustrate the power of using recombination and mutation as means of genetic dissection. It has been possible to determine the fine structure of the gene down to the individual nucleotide. The finest and most complete analyses, of course, are in microorganisms where enormous numbers and strong selective methods make it possible to study very rare events.

These studies also emphasize once again the precision of the crossing-over process. Somehow the exchange takes place with such accuracy that there is not the gain or loss of even a single nucleotide; even at this level, mistakes are rare.

COMPLEMENTATION TESTS IN MAMMALIAN CELL CULTURES

Allelism tests are often not possible in human genetics. The human species is far from ideal for genetic analysis. In particular, if a rare disease is found in two unrelated families and if there are no matings between individuals of the different families (which is almost always the case), there is no way to do a complementation test in the classical way.

Now that human cells can be grown in culture, there is frequently a way around this impasse. Cells of two kinds, grown together in culture, can be made to fuse to form hybrid cells. It is then possible to get the same information as from a genetic experiment with whole organisms. If the hybrid cell, made from parent cells that are both mutant, is normal, then we know that the two types are complementary. If so the two mutants can be presumed to be at different gene loci.

It was mentioned earlier that the recessive disease xeroderma pigmentosum leads to skin tumors when the afflicted person is exposed to sunlight containing ultraviolet radiation. This susceptability to radiation is caused by deficiencies in the enzymes that repair UV damage to the DNA. Two different forms of the disease were shown to be mutants of different genes, since the hybrid cells in culture were able to repair UV damage.

PROBLEMS

1. In Drosophila, white, cherry, and vermilion are all sex-linked recessive eye colors. A white-eyed female crossed with a

vermilion-eyed male produces white-eyed male offspring and wild type red-eyed female offspring. A white-eyed female crossed with a cherry-eyed male produces white-eyed sons and light cherry-eyed daughters. Which genes are alleles (by a complementation criterion)?

2. In rabbits three of the alleles of the c locus are C, c^{ch}, and c. The various genotypes and their phenotypes are:

   C/C, C/c^{ch} C/c - full color
   c^{ch}/c^{ch}, c^{ch}/c - chinchilla
   c/c - white

   A full-colored rabbit is crossed with a white rabbit and produces three chinchilla and four full-colored offspring. What are the genotypes of the parents?

3. What offspring would be expected from the mating C/c by c^{ch}/c?

4. Explain why complementation tests can classify mutant sites as to whether they are in the same cistron or not, but cannot determine their order within a cistron.

5. Do the following data show evidence of position effect? If so, change the numbers so that they don't, or if not, change them so they do.

Genotype	Number of eye facets
$+++/a++$	120
$+++/abc$	60
$a++/abc$	30
$+b+/a+c$	40
abc/abc	20

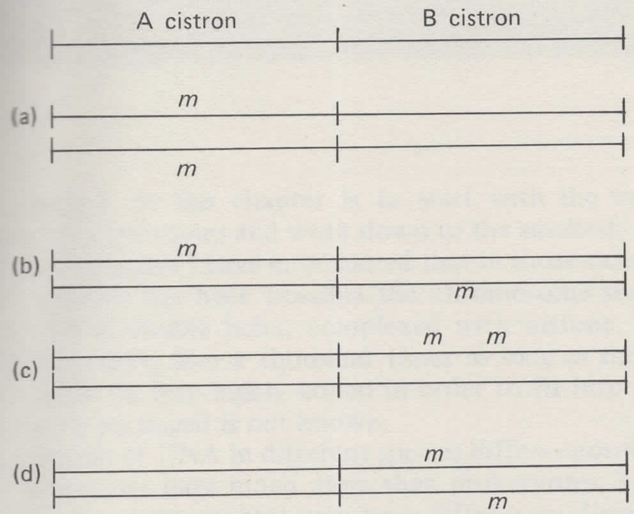
6. A pollen-bearing plant with self-sterility alleles s^1 and s^2 is mated with a female plant of genotype s^2/s^3. What genotypes would be found among the progeny, and in what proportions? What if the female plant were s^1/s^2? s^3/s^4?

7. Mutants a and b are in different cistrons. Will the genotypes $\frac{a\ b}{+\ +}$ and $\frac{a\ +}{+\ b}$ have the same or different phenotypes? If different, which will be normal and which mutant?

8. Wright has pointed out that some alleles are "mixomorphs," in that they behave as hypomorphs in some combinations and antimorphs in others. The allele, pearl (w^p) is such an example. In order of increasing eye pigment, the various combinations are: w/w, w^p/w, w^p/w^p, w^a/w^p, w^a/w, w^a/w^a. Suggest an interpretation.

9. In the bacteriophage Benzer has reported two adjacent cistrons, A and B, each composed of a large number of recons within which separate mutants have been located. Each of the mutants

is recessive in the sense that there is normal phage activity when both normal and mutant phages infect the same bacterial cell. Which of the following combinations would be expected to produce normal and which abnormal (mutant) activity? Assume that everything is normal except for the mutant recon labeled *m*.



CHROMOSOME CHANGES

15

The strategy in this chapter is to start with the more gross changes in chromosomes and work down to the smallest.

In earlier chapters I have emphasized that in those cases where a detailed analysis has been possible the chromosome seems to be one long DNA double helix, complexed with histone. Since the DNA is something like a thousand times as long as the chromosome, it must be very highly coiled in order to fit into the space. Just how it is packaged is not known.

The amount of DNA in different species differs enormously. Of course eukaryotes have much more than prokaryotes, but among the eukaryotes there are still very large differences. Using yeast as a standard, Drosophila has about 7 times as much, the silkworm about 20, the mouse almost 100 times as much, and barley 400 times. Man has slightly more than the mouse. The increase may be due both to an increased number of chromosomes and an increased size.

I shall consider first an extreme, and very useful example of increase in chromosome size.

POLYTENE CHROMOSOMES

The outstanding examples of this are found in the larval salivary glands and other tissues of a number of kinds of flies, including the geneticist's favorite, Drosophila. These chromosomes are many times the size of ordinary chromosomes. They arise by the chromosomes remaining in a permanent interphase or prophase. The replication process goes on, but the replicas lie side by side instead of separating. After about 10 such divisions, there is a bundle of about 1000 strands; so the giant chromosome is really a cable. Such multi-stranded chromosomes are called **polytene chromosomes**. Each strand must itself involve a highly coiled strand of DNA, for the total length of a salivary gland chromosome, long as it is (of the order of 0.1 mm), is much shorter than the DNA strand.

One might wonder how such multi-stranded chromosomes manage to go through mitosis: the answer is that they do not. The salivary glands (and some other larval tissues) consist of cells that enlarge, but do not divide. Then, when the insect metamorphoses into an adult, these tissues degenerate.

These enormous chromosomes have been invaluable for cytological studies. They have characteristic banding patterns that are easily recognized, so the skilled observer can identify not only individual chromosomes but minute regions of chromosomes. It has been possible, as we shall see later, to associate the position of a specific gene with a specific band.

Another advantage for cytological analysis is that the homologous chromosomes are synapsed, just as in the meiotic prophase.

POLYPLOIDY

Cells with more than two sets of chromosomes are known as **polyploids.** The terms diploid and haploid have already been used for cells with two and one set of chromosomes. For specific degrees of polyploidy, such terms as **triploid** (three sets), **tetraploid** (four sets), etc., are used.

Polyploid cells are formed when the process of chromosome reduplication gets out of phase with the process of cell division. If, for example, the chromosomes undergo an extra duplication, a tetraploid cell results; if they undergo two duplications, an octoploid.

Polyploids can be produced in many plants almost at will by the use of various treatments. The usual method is by using colchicine, which interferes with spindle formation.

Polyploids may be classified into two general types:

1. **Autopolyploids.** When more than two sets of chromosomes, all derived from the same species, are present in an individual such an individual is known as an autopolyploid. An autotetraploid would, for example, have four chromosomes of each type. Several common domestic plants (such as orchard grass) are autopolyploids. Polyploids are usually larger and often more vigorous than diploids. As would be expected, triploids, pentaploids, and others with odd numbers of chromosomes have difficulty with synapsis and disjunction at meiosis. As a consequence they usually are sterile.
2. **Allopolyploids.** Individuals with increased chromosome numbers derived from different haploid sets are known as alloploid. If the chromosomes of a first generation hybrid between two different species are doubled, the result would be an allotetraploid. Many inter-species hybrids are sterile because the chromosomes do not synapse, or pair, at meiosis. If the chromosomes are doubled each member has another to pair with and meiosis proceeds normally. Consequently, many allopolyploid hybrids are fertile when the diploid hybrids are sterile.

An example of an allopolyploid is to be found in Raphanobrassica, studied by Karpechenko. The radish (Raphanus) has nine pairs of chromosomes and the cabbage (Brassica) also has nine pairs, but entirely different so that no chromosome corresponds to any in the radish. The hybrid between the two is sterile, but among the hybrids were some fertile plants which turned out to have 36 chromosmes, hence were allopolyploids.

As in most systems of classification, there are intermediate cases which don't fall clearly into either group. Many polyploids are mixed, being alloploid for some chromosomes and autoploid for others.

Polyploidy, though common in plants, is rare in animals. A few polyploid animal species are known and polyploidy has been induced artifically in some, but in the main polyploidy is restricted to plants. However, polyploidy for certain tissues is common in both plants and animals. Mammalian cells grown in tissue culture frequently become polyploid.

In some animals triploids are found. The triploid Drosophila female is quite viable; on the other hand, triploidy in the human leads to embryonic death.

POLYSOMICS

Nondisjunction of a chromosome can lead to a cell with three or one representative of this chromosome in an otherwise diploid cell. Such cells are called **trisomics** and **monosomics**. If two different chromosomes each have triple representation, it is called a double trisomic. If one chromosome is represented four times, it is a tetrasomic; if not represented at all it is a nullosomic, and so on.

Chromosome complements in which every chromosome is represented the same number of times (such as haploids, diploids, and polyploids) are called **euploids**; those in which some chromosomes are represented more times than others are called **aneuploids**.

An important generalization is that euploids are almost always more viable than aneuploids; an organism is able to tolerate quite a change in the total number of chromosomes provided there is the same relative frequency of each type of chromosome as before. It is as if there were certain delicate balancing mechanisms that maintain the same relation if all chromosomes are increased in number equally, but that are upset if one chromosome becomes more frequent than another.

Trisomics have often made it possible to determine which chromosome a set of genes is on. Ordinarily, by breeding tests alone, a geneticist can tell that a group of linked genes are all on the same chromosome, but he doesn't know which chromosome it is. But genes that are present in triple dose show irregular segregation at meiosis, and the geneticist therefore knows that these genes must be on the chromosome that has three representatives.

In many plants trisomics are viable, although they almost always have a characteristic phenotype that departs from normal. In contrast, trisomics in animals usually have drastic effects. We are already familiar with the situation in the human (Chapter 3). Trisomy for even small chromosomes causes gross abnormalities, for some slightly larger chromosomes trisomy leads to abortion, and for the largest chromosomes the death is so early that the embryos are not found. All autosomal monosomic types die before birth.

DEFICIENCIES OR DELETIONS

Whenever a chromosome is broken, the broken ends show a tendency to reattach, either back in the original way or with other broken ends. If the chromosome is broken in two places, the center piece may be left out while the two ends fuse, leading to a chromosome with a **deficiency** or **deletion**.

The principal use of deficiencies has been that of determining, from what happens when certain genes are absent, what they must have done when present. For example, when the part of the chromosome containing the locus of a certain gene is missing, the eye is white instead of red. Hence, we know that the missing part of the chromosome is in some way connected with the formation of red pigment. When a certain portion of the X chromosome of Drosophila is missing, the wing has a notch in the outer border. Some other deficiencies result in shorter bristles. All of these lead to a better understanding of the part played by the normal genes in this region. Most large deficiencies are lethal.

A second use for deficiencies has been in the location of genes on chromosomes. This technique has been used mainly in Drosophila, where the giant salivary chromosomes make precise cytological work possible, and in corn. The order and relative distance apart of many genes in Drosophila have been inferred from crossover data, but by the use of deletions it has been possible to locate a gene in an actual physical spot on the chromosome. For example, in a series of deletions near the "left" end of the X chromosome all were missing the white-eye gene. One particular chromosome band was missing in all these deletions so the locus of the white gene must be somewhere in or near this band. In this manner the precise location of other genes has been determined. In all such studies the gene order previously inferred from linkage studies has been confirmed, though there are some discrepancies in the distances, indicating that crossing-over is more frequent in some parts of the chromosome than in others.

DUPLICATIONS

Just as certain portions of a chromosome may be lost, by the same mechanism parts may be duplicated. Often the duplication of a small fragment of a chromosome may lead to decided phenotypic modifications. The eye character known as Bar is due to a small piece of chromosome being present in duplicate. A normal X chromosome contains a group of four bands (as observed in salivary chromosomes) and in a chromosome causing Bar eye, these bands are duplicated so that an eight-band region results. Sometimes this region is present three times, in which case an intensification of the Bar effect results.

A careful study of salivary chromosomes reveals numerous double or triple bands, whose similarity suggests that they are tiny duplications. The most popular hypothesis is that such duplications arose by mistakes in crossing-over sometime in the history of the organism. This idea is very significant for the theory of evolution for it provides a mechanism whereby the species can increase its number of genes and thereby attain greater complexity. When a gene is duplicated, one of the two is not needed and is therefore free to mutate to new forms. Increasing the number of genes by polysomy or large duplications seems unlikely because these are almost always deleterious, if not lethal.

The first recognized case of position effect is associated with a duplication, the one that results in Bar eyes in Drosophila. If we let *R* stand for the duplicated region, we have the following:

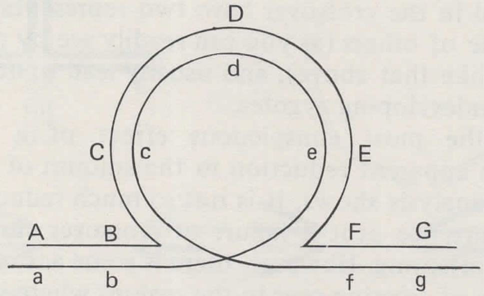
	Chromosome makeup	Number of eye facets
1. Normal	*R/R*	750
2. Heterozygous Bar	*RR/R*	360
3. Homozygous Bar	*RR/RR*	68
4. Heterozygous Double Bar	*RRR/R*	45
5. Homozygous Double Bar	*RRR/RRR*	25

The position effect is indicated by the difference in the size of the eye in genotypes 3 and 4. Each has the same total chromosome material, but having three blocks on one chromosome and one on the homologue gives a smaller eye than having two blocks on each.

INVERSIONS

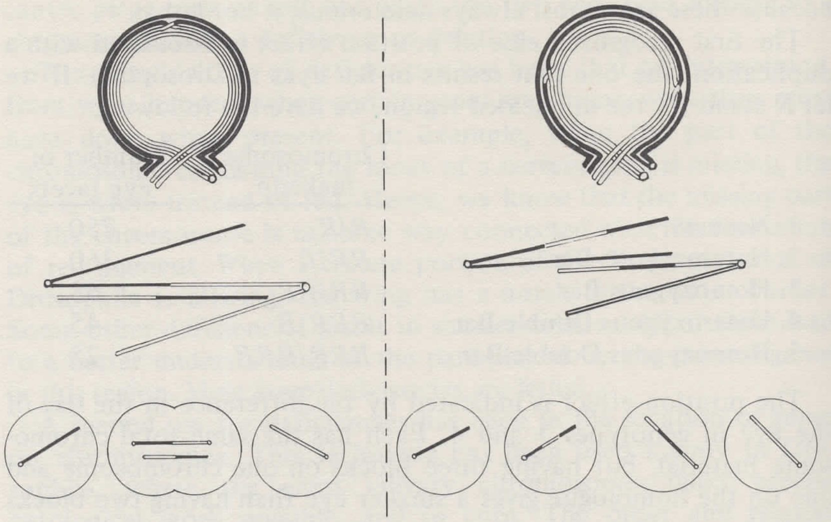
Sometimes, as a result of a doubly broken chromosome, the center piece is inverted. For example, if a chromosome with the gene order *a b c d e f g h* is broken between *b* and *c* and between *f* and *g* and the center portion turned through 180°, the resulting gene order is *a b f e d c g h.*

One thing that the study of inversions has shown is the exactness of synapsis. Whenever it is mechanically possible the chromosomes synapse such that corresponding genes lie opposite one another. For example, synapsis with an inversion often leads to a loop formation like this (the original gene sequence is indicated with capital letters, the inverted chromosome with lowercase letters):



This shows how remarkably precise the process of synapsis is. More complex rearrangements, such as inversions within an inversion, lead to still more complex patterns, but the general principle of having corresponding genes opposite each other still holds.

An important consequence of a heterozygous inversion is its effect on crossing-over. If we diagram the synaptic configuration at a four-strand stage showing a crossover between the second and third strands we have something like the diagram on the left.



At the top we see the four chromatids just after the exchange event. Below is shown the next anaphase (Anaphase I). The two recombinant chromosomes have either two centromeres or none. The bottom diagram shows the four products of meiosis. The only normal eggs or sperms are the two that contain chromosome strands not involved in the exchange. Hence the result of inversions is to prevent the recovery of crossover strands. On the other hand, as shown in the diagram on the right, a double exchange involving the same two strands has no such effect.

Notice that a similar consequence occurs when the centromeres are within the inverted region. This time there are no dicentric or acentric chromosome strands (i.e., strands with two or no centromeres), but aneuploid gametes are produced. The chromosomes involved in the crossover have two representatives of some genes and none of others (as you can readily see by putting letters on a diagram like that above), and usually lead to nonfunctioning gametes or nondeveloping zygotes.

Therefore the most conspicuous effect of a heterozygous inversion is an apparent reduction in the amount of crossing-over. As the above analysis shows, it is not so much reduction in actual crossover occurrence as it is failure of crossover chromosomes to lead to viable offspring. However, there is some actual reduction in the occurrence of crossing-over in the regions where synapsis is not perfect.

In the early history of Drosophila genetics **crossover suppressor factors** were found. It must have seemed quite surprising to these early workers to find a "gene" that suppressed crossovers when heterozygous, but not when homozygous, and which suppressed single crossovers, but had no apparent affect on double crossovers! The first person to realize that these paradoxical findings could be readily explained by an inversion in gene order was A. H. Sturtevant.

A paradoxical observation is that an inversion in one pair of chromosomes which reduces the amount of crossing-over in this pair, will actually enhance the amount of crossing-over in another pair of chromosomes, or even in a distant part of the same chromosome pair. So far there isn't any satisfactory explanation of how this occurs.

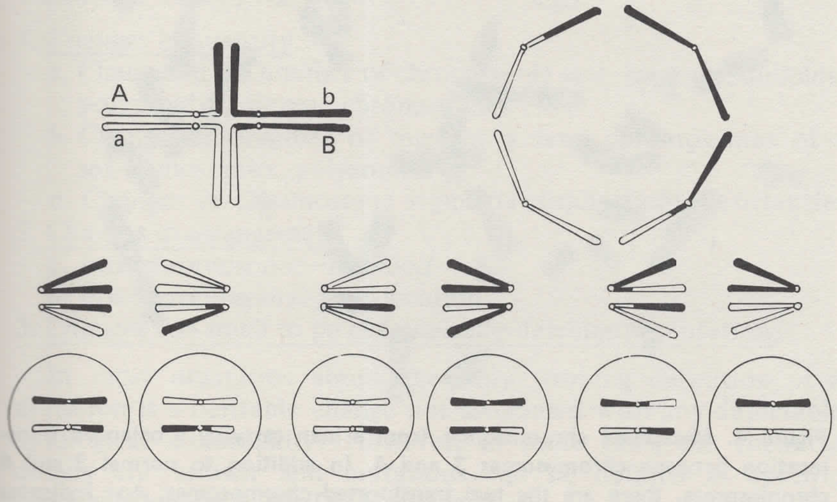
TRANSLOCATIONS

When breaks occur in two different chromosomes and there is an exchange of parts, the result is called a reciprocal **translocation.** If the parts attach in such a way that one chromosome has two centromeres and the other none, the effect is usually the death of this cell or of descendant cells. But if they attach in such a way that each piece has a centromere, the rearrangement may persist.

Translocations are usually detected by a change in linkage relationships. If genes known to have been on separate chromosomes start behaving as if linked, it indicates a translocation between the two chromosomes.

At meiosis with a heterozygous translocation, an X shaped configuration is necessary in order that all genes can lie close to their alleles. Since many of the gametes that are formed from an individual heterozygous for a translocation are unbalanced, either the gametes do not function (especially pollen in plants) or the zygotes do not develop (as usually happens in animals, where aneuploid sperms and eggs function perfectly normally). But in either case there is a failure of offspring to develop, so a translocation typically results in partial sterility. A semi-sterility that is regularly transmitted to half the offspring usually has this explanation.

This is illustrated in the following drawing.



The upper left configuration shows the chromosomes at the time of synapsis. Genes *A* and *B* are on nonhomologous chromosomes that were involved in the exchange. At the upper right the chromosomes are starting to separate. (The drawing is simplified in not showing both members of a chromatid pair.) The six different kinds of gametes are shown at the bottom. Notice that only the right two are balanced. Notice also that these two contain either *A* and *B* (left) or *a* and *b* (right). Thus the genes *A* and *B* behave as if linked as a result of the translocation.

Although mongolism is usually caused by nondisjunction during meiosis of a normal woman and thus rarely occurs twice in the same family, there are occasionally families with several children having mongolism. This is readily understood as the result of a translocation between the mongolism chromosome and some other chromosome. A parent carrying a balanced translocation would produce some unbalanced gametes, leading to duplications and deficiencies. The deficiencies lead usually to embryonic death, but the duplications (if they include a substantial part of the mongolism chromosome) produce mongoloid children.

The identification of the chromosomes is of great value in such cases, for a skillful cytologist can often identify normal persons who carry a translocation and who, therefore, have a high risk of having abnormal children and prenatal deaths.

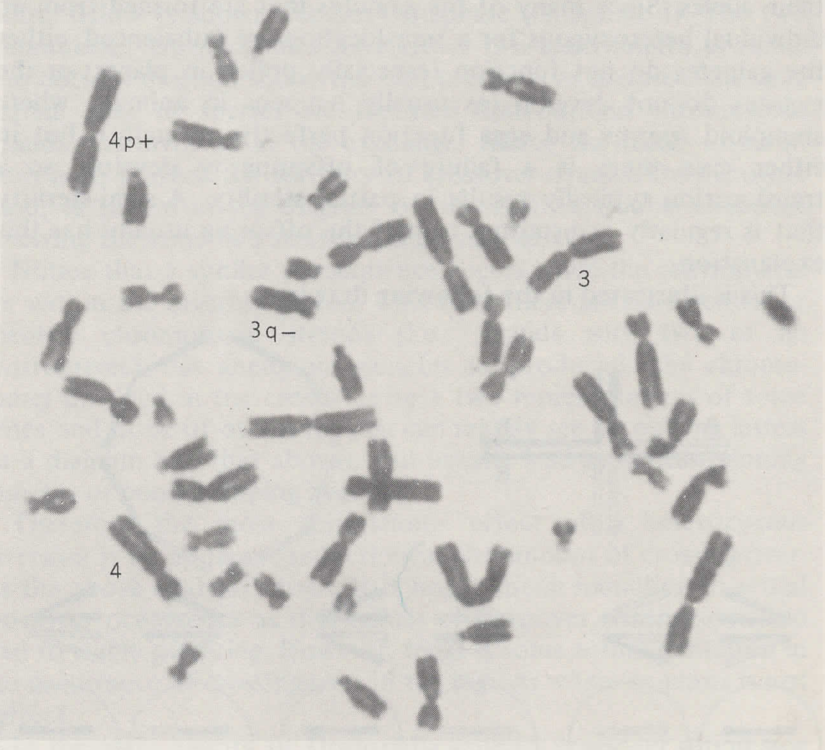


Figure 4. Metaphase chromosomes from a man carrying a balanced translocation between chromosomes 3 and 4. In addition to normal 3 and 4 chromosomes, there are the two translocated chromosomes: 4p+ indicates that material has been added to the short arm of 4, and 3q- indicates material removed from the long arm of 3. Although the man is normal he is in effect partially sterile because of the loss of embryos with unbalanced combinations.

AGENTS THAT BREAK CHROMOSOMES

The most effective chromosome-breaking treatment is high energy radiation (X rays, gamma rays, neutrons, alpha and beta particles) that causes ionization. Less effective are a number of chemicals, especially mustard gas. In general, agents that produce mutations also break chromosomes, though not with equal effect.

V-TYPE POSITION EFFECTS

For genetic analysis it is important to distinguish two kinds of chromatin, **euchromatin** and **heterochromatin.** Heterochromatin is frequently found near the spindle attachment, but may be elsewhere, and is characterized by being more condensed and darkly staining during the prophase. Most of the genes are in the euchromatin, the heterochromatin being largely inert in this respect. For example, the Y chromosome of Drosophila is heterochromatic and has very few genes.

When a translocation or an inversion moves a gene that was previously in the euchromatin near to the heterochromatin, the function of the gene is often altered. For example, moving the white-eye gene from its normal position, to a position in or near the heterochromatin of any of the chromosomes results in an eye that is variegated or mottled in color. Since the total gene makeup is not changed, but merely the arrangement of the genes, this is another example of position effect. Heterochromatin-caused position effects are called V-type (for variegated) since they typically result in a mottled phenotype.

5

It is interesting to note that the few genes normally located in the heterochromatin show a variegated phenotype when they are moved to the euchromatin. So it appears that each gene is adapted to function in a particular type of chromatin, and its function is upset if it is moved to the other type.

CLASSIFICATION OF CHROMOSOME CHANGES

It may be convenient to summarize the types of chromosome changes by classifying them.

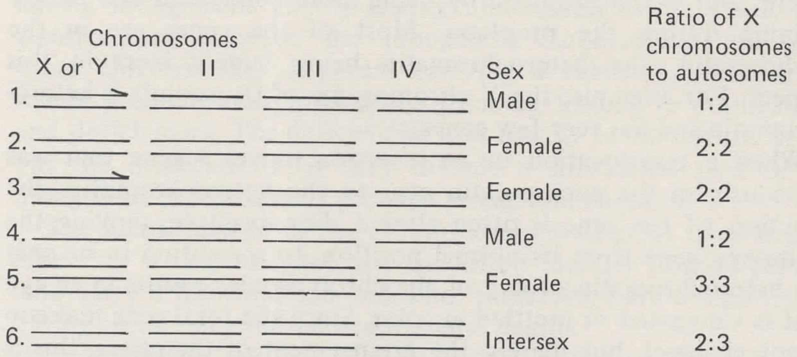
1. Changes in quantity
    a. Changes in the number of chromosome sets–haploids, diploids, polyploids, polytene chromosomes
    b. Changes in number of one (or a few) chromosomes of a set–monosomics, polysomics
    c. Changes in chromosome segments–duplications, deficiencies
2. Changes in arrangement
    a. Intrachromosomal–inversion
    b. Interchromosomal–translocation
3. Changes too small to be cytologically detectable–mutation

In most organisms about the only working definition of a mutation is a heritable change not associated with any detectable cytological abnormality. However, in some cases the chemical details are known and mutations can be defined as specific chemical changes in the DNA molecule.

AN EXAMPLE OF GENIC BALANCE: SEX DETERMINATION IN DROSOPHILA

An excellent illustration of the fact, emphasized earlier, that normal development depends on a balance between the genes on different chromosomes is provided by the sex determining mechanism in Drosophila. Recall from Chapter 3 that the Y chromosome has nothing to do with sex determination.

The following diagrams show that sex is determined, not by the absolute number of X chromosomes, but by the number of X's relative to the number of sets of autosomes.



Apparently there are genes on the X chromosome that tend toward femaleness that are balanced against autosomal genes tending toward maleness. It has been shown by study of duplications that there are many genes along the X chromosome that are concerned with sex determination, not just a single gene.

Another example of genic balance sex determination is illustrated by Melandrium, a higher plant. This time the balance is not between the X and autosomes, but between the X and Y chromosomes. XX, XXX, and XXXX plants are all female regardless of the number of sets of autosomes; XY, XXY, and XXXY are males. The fact that plants with chromosome makeup XXXXY are intersexes shows that one Y chromosome is roughly equal to 4 X's in sex determining potency.

CHROMOSOME MAPPING IN HUMAN CELL CULTURES

Classical gene mapping by recombination is very difficult in the human species. Small sibships, the inability to control matings, long life cycle, and the rarity of many of the genes combine to make the human a very poor choice for gene mapping. If it were not for our special interest in our own species, we would have given up trying to map human chromosomes long ago. As it is, by taking advantage of fortuitous matings that were informative about linkage relations and by some very clever statistical procedures, human geneticists have managed to discover a few linkages. But the human gene map has been in a very primitive stage compared to that of *E. coli,* Drosophila, corn, or the mouse.

The recent discovery that cells from different species, often very far removed phylogenetically, can be made to hybridize in cell cultures has changed all this. Some viruses, especially a strain known as the Sendai virus, somehow promote such cell fusions. Cells with human and mouse chromosomes together grow very well in culture. However, the human chromosomes do not divide as rapidly as the mouse chromosomes, so that after a number of cell divisions the nuclei contain mostly mouse chromosomes.

A cytologist can take advantage of this to find the chromosomal location of genes, provided the genes produce some phenotype that can be recognized in cell culture. Fortunately many enzymes and other proteins can be identified in cultured cells by appropriate chemical tests. Also fortunately, recent staining techniques mentioned in Chapter 3 make it possible to recognize individual chromosomes easily. Therefore the cytologist can observe a cell culture and note that a particular human chromosome is missing. The cytologist can then analyze the culture and see what gene function is also missing, and thereby identify which chromosome this gene is on.

Similarly, since chromosomes often get broken and rearranged in cell cultures so that parts of chromosomes get lost, it is often possible to locate the gene in a certain chromosome region. Then by combining knowledge from overlapping regions the location can be determined with increased precision.

THE NUMBER OF GENES IN DROSOPHILA

The number of bands in the salivary chromosomes of Drosophila is about 5000. It was suggested very early after the discovery of these chromosomes that this might represent the number of genes; now it looks as if this is probably correct.

For many years B. Judd has collected recessive lethal and visible mutations that are located within a 16-band deletion near the left end of the Drosophila X chromosome. Several hundred were obtained. These were then tested in pair-wise combinations for complementation. If the lethals are in the same cistron the combination of the two lethals on homologous chromosomes is lethal; if they are in different cistrons the combination produces a normal phenotype. The striking result is that this region turned out to have 16 cistrons. So it looks as if the number of cistrons is the same as the number of bands in the salivary gland chromosomes. Therefore, it we equate the gene with the cistron, there are about 5000 genes (this is the haploid number) in Drosophila.

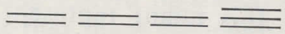
Similar, but less extensive, studies by others on other parts of the chromosomes have yielded results consistent with this.

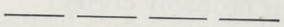
This raises a problem. Five thousand cistrons, each 1000 nucleotides long, is 5 million. But the number of nucleotides in Drosophila DNA is about 1.4×10^8, about 30 times this many. It is clear that most of the Drosophila DNA is doing something other than coding for proteins, while in *E. coli* most of the DNA does code for proteins. Perhaps the rest of the DNA in higher organisms is needed for regulation, timing, and other activities required to coordinate the development of a multicellular species.

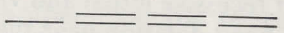
The number of genes (cistrons) in mammals is unknown. As has been emphasized before, the amount of DNA is much larger than in Drosophila. But there is no reason to think that the number of genes has increased in proportion. The best guess is that the number is not much larger than that of Drosophila.

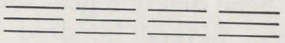
PROBLEMS

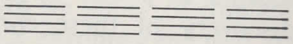
1. A Drosophila is heterozygous for a reciprocal translocation between the second and third chromosomes, each break having occurred near the spindle attachment (which for each of these chromosomes is near the center). Draw a diagram showing how these chromosomes would synapse at meiosis.
2. A male Drosophila heterozygous for the translocation mentioned in problem 1 and with the recessive genes brown eye (*bw*) and ebony body (*e*) on the nontranslocated second and third chromosomes respectively, is crossed to a female with normal chromosomes and homozygous for the genes *bw* and *e*. What offspring would be expected? (Any zygotes with an extra chromosome arm or deficient for one do not survive.)
3. Fewer recessive phenotypes appear among the offspring of self-pollinated tetraploids than among those of self-pollinated diploids. Why?
4. Assign the correct names from the following list to the figures below: trisomic, tetrasomic, tetraploid, double trisomic, triploid, haploid, nullosomic, monosomic.

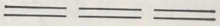
a. 

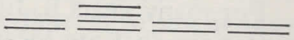
b. 

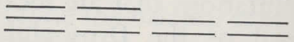
c. 

d. 

e. 

f. 

g. 

h. 

5. Persons with mongolism are feebleminded and usually do not reproduce. In the event that a mongoloid does have children, what are their chances of mongolism?
6. A. H. Sturtevant found a gynandromorph Drosophila that began as a zygote with the following X chromosomes:

$$\frac{s+e+c\ \ v\ \ g\ \ f}{+w+r+++f}$$

The male parts of the fly showed the mutants *s*, *e*, *c*, *g*, and *f*. The female parts showed only *f*.

a. How was this gynandromorph formed?

b. Which mutants are autonomous, which nonautonomous, and for which does this fly offer no evidence?

7. What kinds of chromosomes would be produced after a four-strand double exchange within an inversion? Assume (a) that the centromere is outside the inverted region, and (b) it is inside.

MUTATION

16

GERMINAL AND SOMATIC MUTATION

Mutations can occur in various kinds of cells, but most of them probably have very little effect on the individual and are ordinarily undetected. If a mutation occurs in a somatic cell, it is transmitted to all the descendant cells. Bud sports in fruit trees are probably somatic mutations, and are of practical importance since they can be propagated by grafts. Many commercial varieties of fruits originated in this way.

But the mutations that are of the most consequence from the standpoint of heredity are those that occur in germ cells, for these can be passed on to the next generation. Mutation rates are very low for individual genes, but there are so many genes that the likelihood of a mutation in one or another of them is fairly high.

MUTATION RATES FOR INDIVIDUAL GENE LOCI

In most higher organisms, plant and animal, the most commonly observed mutation rates are around 1 per 100,000 genes per generation. Values 10 times higher or lower than this are often found. Measurement of human mutation rates is very uncertain, but the estimates made place the average rate in this same range.

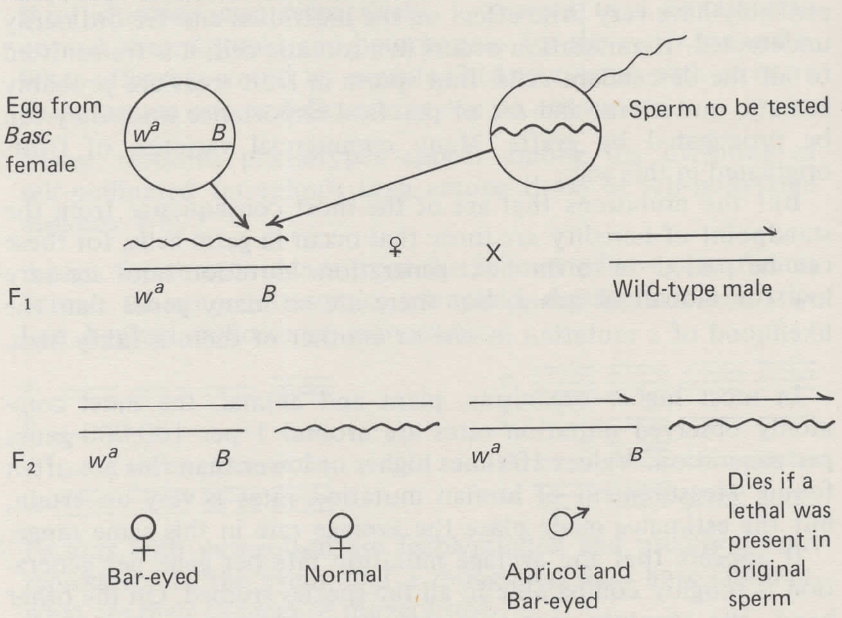
It appears that the average mutation rate per gene per generation is roughly comparable in all the species studied. On the other hand, the absolute mutation rate (the rate per year) must be greatly different. Under favorable circumstances Drosophila may have 20–30 generations per year, so there may be as many as 1000 Drosophila generations per human generation. So there must have been evolutionary adjustment of the mutation rate to conform to the life cycle.

MUTATION RATES FOR WHOLE CHROMOSOMES

There is an advantage in measuring the total rate for all genes for two reasons. One is that in order to assess the total effect of mutation on the organism, we need to know the total mutation rate. The other is that if we want to determine the effect of various treatments on mutation rates, it is impractical to study individual genes; their mutation rate is so low that only an extremely large increase would be detectable.

The precise methods of mutation measurement that are now widely used in Drosophila research are all similar in principle to a method invented by H. J. Muller in the 1920's. His first method, the ClB, is now mainly of historical interest, so we shall present the current method, which is the same except for some minor improvements in the choice of strains of Drosophila to use.

Muller's method detects whether there is a lethal mutation in the X chromosome of a particular sperm. This must have occurred during the lifetime of the male that produced the sperm, else the embryo from which he developed would have died. We indicate the chromosome to be tested for the presence of a lethal by a wavy line. The females come from a stock known as Muller-5, or *Basc.* The X chromosome has two mutants w^a (apricot eye color) and *B* (Bar eye shape) and has a complex inversion within an inversion that permits practically no recoverable crossovers. (The name *Basc* is derived from the initials for Bar, apricot, and scute-8, the name for one of the inversions.)



If the F_1 females are mated individually in separate containers so that all the progeny in a single container come from a single female, then all males that are not apricot and Bar-eyed have the same X chromosome as the original sperm to be tested. Thus, if no such males are found, there must have been a lethal in the original sperm. Or if the sperm carried a visible mutation, all the non-apricot, non-Bar males will show the mutant phenotype.

The result of numerous such studies shows that about two lethal mutations occur per thousand X chromosomes per generation. The X chromosome makes up roughly 1/5 of the chromosome material in a sperm, so that about one sperm in a hundred carries a new lethal.

Drosophila studies suggest that mutations causing a harmful, but not completely lethal, effect occur about 10 times as frequently as lethals, so the total rate of mutation must be such that roughly one gamete in 10 carries a new mutation of some sort.

FACTORS AFFECTING MUTATION RATES

The existence of precise procedures like the C1B and Muller-5 methods has made it possible to study the influence of various agents on the mutation rate. The results of hundreds of studies may be briefly summarized as follows:

1. Temperature. Within the range of temperatures that a fly can tolerate, the mutation rate increases with an increase in temperature. A 10° temperature rise results in a four- or five-fold increase in the number of mutations produced.
2. Radiation. In 1927 Muller made the discovery that X rays greatly increase the rate of mutation. He X-rayed normal males and by the C1B procedure showed that the sperm cells produced by these males had many more lethal mutations than those of the unirradiated control flies. The number of mutations produced is directly proportional to the X-ray dose.

   One might wonder if, since radiations are known to produce mutations, spontaneous mutations might not be due to natural radiations. However, this turns out to be completely untenable, at least in Drosophila and mouse. It would require about 50 Roentgens to produce as many mutations as occur spontaneously in a single Drosophila generation; yet during this time a Drosophila receives only about .004 Roentgens. (A Roentgen is a unit of radiation amount, named after the discover of X rays, and frequently abbreviated R. It is equivalent to about two ionizations per cubic micron of protoplasm.) Similar considerations apply to the mouse; natural radiations account for less than 0.1 percent of spontaneous mutations. The rest must be due to other causes. On the other hand, man, with his much longer life cycle during which he absorbs correspondingly more radiation, probably has a larger fraction of his spontaneous mutations due to radiation. The fraction is unknown, since no data are available on the rate of mutation from radiation in man, but based on mouse data is estimated at about 10%. In addition to X rays, radiations of a wide range of wave length produce the same result. Gamma and Beta radiation, neutrons, and ultraviolet have been used successfully.
3. Chemicals. Until 1945 it appeared as if the mutation process could not be influenced by external chemical treatment, but in that year Charlotte Auerbach announced the increase in mutation rate with mustard gas. Since then many compounds (formaldehyde, some peroxides, phenol, certain carcinogens) have been shown to be mutagens.

The most precise measurements of mutation rates and the effect of chemical mutagens have been in microorganisms. Particularly neat experiments were done by Dr. Aaron Novick using a device called the *chemostat.* This is a glass container in which the culture medium is being continually added while an overflow valve drains off the excess. As a consequence, the bacterial culture soon comes to a steady state in which the total number of cells remains constant, and the rate of cell multiplication is exactly balanced by loss of cells through the overflow. The rate of cell division can be completely determined by the rate at which the medium is added.

In this system new mutations accumulate linearly with time and the mutation rate can be very accurately determined by taking samples at various times and assaying for the number of mutant cells. A chemical treatment that changes the mutation rate will cause a change in the slope of the line when mutations are plotted against time.

Novick tested a number of compounds related to nucleic acid, and found that several of them increased the mutation rate. He also discovered for the first time (because this was the first time anyone had devised a system sensitive enough to detect it) antimutagens, i.e., substances that decrease the spontaneous mutation rate. Two substances, guanosine and adenosine, substantially reduced the normal rate. Thus it appears that the "spontaneous" mutation rate is determined by the balance of mutagenic and antimutagenic substances in the cell.

THE IMPORTANCE OF MUTATIONS

The ability to mutate is a unique property of living organisms. It is also a property that makes life and evolution possible, for without mutation selection would have no variations on which to act.

Since mutations are random occurrences, in any well-adapted organism they are much more likely to be harmful than beneficial. Having to put up with a large number of harmful mutations is a price that must be paid by the population for the privilege of evolution. For it is only by the occurrence of a large number of mutations, most of them harmful, that the rare beneficial mutation or combination of mutations can have an opportunity to occur.

Since the overwhelming majority of mutations are harmful, they constitute a serious problem for man. Nobody knows just what proportion of our disease is ultimately attributable to mutation, but perhaps a sizable fraction is. Most people carry several hidden recessive mutants which, if they were to become homozygous, would have a lethal or at least a deleterious effect. One of the most serious problems raised by the increasing use of chemicals in our lives is that some of them may be mutagens. The mutations induced in this generation may not be manifest for many years in the future, but eventually they will become expressed and do their harm. It is obvious that every reasonable precaution should be taken to protect the germ cells from radiation and mutagenic chemicals.

THE MOLECULAR BASIS FOR MUTATION

Most spontaneous mutations, as observed in sensitive systems where fine distinctions can be made, seem to involve a single nucleotide substitution. The mutations that have been incorporated as evolutionary changes in hemoglobin, cytochrome and other proteins where exact amino acid sequences have been determined for many related species, are also mainly of this type.

Mutations of a single base are usually classified into two types: **transitions** are changes of one purine into another, or one pyrimidine into another; **transversions** are changes from a purine to a pyrimidine, or vice versa.

It would be expected that chemicals that affect purines and pyrimidines would act as mutagens. An example is nitrous acid, which is known to change cytosine to uracil. Uracil can form hydrogen bonds with adenine. Thus, treatment with nitrous acid would be expected to change a C-G pair to a T-A pair, through the sequence C-G, U-A, T-A. Likewise, nitrous acid changes adenine to hypoxanthine, which pairs with cytosine, and this leads to the change from A-T to G-C. Thus both kinds of transitions $A \rightleftharpoons G$ and $C \rightleftharpoons T$ can be produced.

Another class of mutagens is base analogs—substances that are structurally similar to the regular bases and which are sometimes mistakenly incorporated into the DNA. For example, bromouracil is quite similar to thymine and can be substituted for it in DNA synthesis. However, unlike thymine, bromouracil sometimes pairs with guanine, thereby producing a transition mutation.

A third class of mutagens is exemplified by proflavine. This causes the addition or subtraction of a single base and thus produces tiny duplications and deletions.

Radiation with ultraviolet light produces mutations by a still different mechanism. It causes thymine molecules that are close together to join tightly as a dimer. This then somehow leads to errors in DNA copying, perhaps by causing stresses in the molecule. Most strains of bacteria have enzymes that are able to remove the thymine dimer and replace it with normal components, thus repairing the damage done by the ultraviolet. Some mutants lack an enzyme and are easily detected by their extreme sensitivity to ultraviolet.

X rays and other ion-producing radiations probably break the DNA strand. One kind of evidence is the efficiency with which these radiations break chromosomes. Perhaps the neatest evidence comes from some virus DNA's that are only single stranded. These are extremely sensitive to killing by radiation. With normal double stranded DNA when one strand is broken the other seems to act as a splint and holds it together long enough for repair to occur. A single stranded DNA, on the other hand, has no such protective mechanism and falls apart when broken; hence its much greater sensitivity to radiation.

Spontaneous mutations in *E. coli* are mostly transitions. One possible mechanism is a tautomeric shift in adenine, suggested originally by Watson and Crick. By shifting a hydrogen from the amino group to the nitrogen in the ring the adenine is changed so that it can now pair with cytosine. Such a state has a low probability, but when it happens to be in this state at the right time the result would be the incorporation of a C opposite it instead of the usual T, thus leading to the A-T to G-C transition.

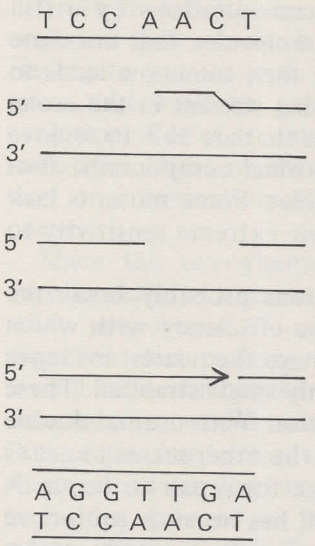
An interesting proof of the genetic control of mutation and an explanation of a mutator gene comes from studies on *E. coli.* This gene somehow greatly increases the probability of A-T to C-G transversions throughout the genome. Thus the mutation rate is generally increased. This makes it obvious that the rate of mutation in general can be increased or decreased by natural selection for or against mutator genes.

MUTATION REPAIR SYSTEMS

There are now several known ways in which abnormal DNA can be repaired. One of the best known is a system for excising the thymine dimers and replacing them with normal bases. It uses four enzymes:

1. **Endonuclease**– this enzyme cuts one strand of the DNA.
2. **Exonuclease**–this removes bases from an exposed end of a single strand.
3. **Polymerase**– as we learned before, this synthesizes a new strand from a template and works in the direction from 5′ to 3′.
4. **Ligase**–joins pieces of DNA together.

The sequence of changes for removing the dimer can then proceed this way:



Irregularity caused by the dimer

Nicking of one strand by endonuclease

Removal of nucleotides from the free end by exonuclease

Synthesis of a new strand by polymerase

Tying together by ligase

It is of interest that the human disease, xeroderma pigmentosum, which was mentioned in Chapter 10 as an example of environmental effect on gene action, can now be explained in molecular terms. Persons with this condition lack one of the repair enzymes and are therefore unable to repair the damage done by the ultraviolet in sunlight. The unrepaired DNA damage somehow leads to skin cancers.

I mentioned earlier that the molecular details of recombination are not understood. Processes such as the one just described could possibly be instrumental in the process of breakage and rejoining

E. coli. This
A-T to C-G
tion rate is
he rate of
by natural

l DNA can
xcising the
t uses four

A.
of a single

ew strand
3′.

can then

er

nuclease

he free end

olymerase

igmento-
mple of
ained in
he repair
done by
omehow

bination
ed could
ejoining

---

involved in normal crossing-over. Two observations favor this. One is that a small amount of DNA synthesis is known to occur during the prophase of meiosis, at the time when crossing-over is thought to occur. The other is that bacterial strains that are deficient in recombination ("rec-less") are often unusually susceptible to mutational damage, and have been shown to be deficient in repair systems. So it is very likely that repair of DNA damage and recombination depend on the same fundamental processes.

USE OF CELL CULTURES TO STUDY HUMAN MUTATION

The great precision and efficiency with which bacterial mutants can be studied is mainly due to the enormous numbers that can be grown. This not only reduces statistical errors but also makes possible, through the use of selective growth media, the isolation of a mutant cell amongst hundreds of millions of normal cells. Some progress has been made in applying these ideas to the study of mammalian cells in culture. This offers the possibility for a precise study of human mutation rates and the factors influencing them.

One approach is to study mutations affecting enzymes that convert purines and pyrimidines or their nucleosides to the corresponding nucleotides, which then can be incorporated into nucleic acids. These "salvage" enzymes are not essential under ordinary conditions since nucleotides can be synthesized in other ways. Therefore mutant cells that are deficient for the salvage enzymes can survive in cultures, and this offers a way of selecting for rare mutants and against normal cells. The general approach is to add a purine or pyrimidine analogue to the medium. Normal cells convert this into a nucleotide which is incorporated into nucleic acid. The nucleic acid with abnormal bases is lethal to the cell. Hence normal cells with the salvage enzyme die, but those that lack the enzyme survive by not being able to make the abnormal nucleic acid.

Robert V. Miller

For example, the enzyme thymidine kinase converts thymidine into its nucleotide which can be incorporated into DNA. Mutant cells can be selected in a culture medium containing bromodeoxyuridine (BUDR), which behaves very much like thymidine. Normal cells incorporate this into their DNA and are killed, while the rare kinase-deficient mutants form colonies of cells. Systems such as this make possible the quantitative study of mutation rates in man. In particular, this kind of study will be of benefit in studying the effects of possible mutagens in the environment, something that would be totally impractical from direct observation of humans. Of course, precise studies can be made and are being made with bacterial cells and with Drosophila, but then you are not sure that the results are applicable to humans.

Another salvage enzyme gene that has been used for human mutation rate study is the one for hypoxanthine-guanine phosphoribosyltransferase (HGPRT). This converts the normal purines guanine and hypoxanthine to their nucleotides. The enzyme also makes toxic nucleotides from the purine analogues azoguanine and

6-mercaptopurine (which is the basis for the use of 6-mercaptopurine in cancer chemotherapy). Thus the mutant cells deficient for the salvage enzyme can be selected and the mutation rate measured with precision. A suspected mutagen can be added to the culture to see if it increases the mutation rate.

This mutant is important for another reason. Although the enzyme deficiency doesn't disturb ordinary cell growth, it disrupts purine metabolism in a way that causes a tragic human disease, the Lesch-Nyhan syndrome mentioned in Chapter 6. Being recessive and X-linked, the disease is confined to males. It is characterized by hyperactivity, mental deficiency, and compulsive self-injury. The deep chemical and cellular understanding makes possible the detection of carrier women and genetic counseling; it may some day lead to a cure.

PROBLEMS

1. Why is so little known about recessive mutation rates for recessive genes in the human species, whereas some rates for dominant mutants are known? Why is the situation different in Drosophila and the mouse?
2. Why is it necessary to have crossover suppressors in the Muller-5 X chromosome?
3. Would you expect mutants produced by nitrous acid to be reversible by nitrous acid?
4. A cell is treated with nitrous acid, and a molecule of cytosine changed to uracil. Two generations later there will be four double helices. What will be their base compositions at this site?
5. Would you expect mutants produced by bromouracil to be reverted by proflavine?
6. Many X-ray-induced recessive lethal mutations never mutate back to normal. Why?
7. If you find a bacterium that is unusually susceptible to UV radiation, what might you suspect as an explanation?
8. If you find a virus that is unusually sensitive to X-radiation, what would you suspect about the virus?
9. Spontaneous mutations are often transversions. Would these be revertible by nitrous acid?

se of 6-mercapto-
nt cells deficient
he mutation rate
can be added to
n. Although the
rowth, it disrupts
uman disease, the
. Being recessive
is characterized
ulsive self-injury.
akes possible the
ng; it may some

tation rates for
s some rates for
uation different

s in the Muller-5

rous acid to be

cule of cytosine
re will be four
ions at this site?

mouracil to be

s never mutate

ceptible to UV
ion?

to X-radiation,

s. Would these

17

CYTOPLASMIC HEREDITY

THE CYTOPLASM

We have seen that the genetic material resides in the chromosomes, but that many cell activities—in particular, protein synthesis—occur in the cytoplasm. The cytoplasm is also the site of many other enzymatic activities.

In addition to such functions, which are largely controlled by nuclear genes, the cytoplasm is also the carrier of some hereditary factors.

CHARACTERS CARRIED BY PLASTIDS

All green plants have chlorophyll-bearing plastids in the cytoplasm. These tiny bodies remain in the cytoplasm and are passed on more or less at random from cell to cell during the mitotic divisions. Since the sperm nucleus coming from the pollen grain carries very little, if any, cytoplasm with it, the plastids are transmitted from generation to generation by the egg.

In several kinds of plants albino and simi-albino chlorophyll-deficient variants are known. Sometimes a cross produces results like this

green female x albino male → green offspring
albino female x green male → albino offspring

The fact that the offspring are always like the maternal parent suggests that the hereditary transmission is by way of the cytoplasm rather than the nucleus. Various other tests have shown this to be the case. This plastid character is perpetuated independently of the genes. Nevertheless it should be pointed out that there are many more genes known to influence the plastids than there are hereditary determiners carried by the plastids themselves.

CYTOPLASMIC INHERITANCE

An example of cytoplasmic inheritance is male sterility in corn. The sterility is due to the failure of the pollen grain to develop properly. Plants of this type pollinated with pollen from any strain produce pollen-sterile F_1 hybrids. Crosses made with chromosomes marked by known mutant genes rule out the possibility that the sterility factor is associated with any chromosome. It depends on a self-reproducing cytoplasmic factor.

In domestic corn strains there are also nuclear genes that can override the cytoplasmic male-sterility factor. The combination of maternally inherited male sterility and nuclear genes that can restore fertility has been exceedingly useful in the commercial production of hybrid corn. This enables the breeder to prevent unwanted pollen for some crosses but restore the fertility for other crosses involving some of the same strains.

The great reliance on a single cytoplasmic factor for much of the commercial corn in the United States led to a near-disaster in 1970. It turned out that this particular cytoplasm was also especially sensitive to a mutant strain of a fungus causing leaf blight. Now more emphasis is given to having a variety of strains available as a hedge.

KAPPA IN PARAMECIUM

Some strains of the protozoan, Paramecium, have hereditary cytoplasmic particles. These particles are unable to reproduce except in the presence of a certain gene. The particle, **kappa,** is responsible for producing a potent antibiotic that destroys any Paramecium not containing kappa. By taking advantage of the nuclear and cytoplasmic behavior of Paramecium during conjugation and autogamy, T. Sonneborn was able to demonstrate that kappa will not reproduce itself unless the gene *K* is present. In the genotype *kk* the particle will not reproduce and is finally diluted out in successive cell divisions. Gene *K* cannot produce kappa, but if only a single kappa particle is put into the cytoplasm of an animal of genotype *KK* or *Kk,* the particle will multiply.

After all this was worked out it was discovered that the particles are really bacteria that live in the cytoplasm of the protozoa and that somehow require some product of the *K* gene for reproduction.

CYTOPLASMIC PARTICLES AND DNA

Particles that replicate autonomously in the cytoplasm have all been shown to have nucleic acid, usually DNA, in very minute amounts compared to the chromosomes. Plastids contain DNA and protein synthesizing machinery. So these are very much like small organisms in the cytoplasm. The similarity of structure with some of the most primitive unicellular plants suggests that they may have originated from infection of a cell by such a tiny parasite.

The mitochondria also contain DNA. Here, as is often the case, some of the proteins are produced by cytoplasmic DNA and some by nuclear DNA. They, too, may have originated from a microbial parasite in the ancient past.

CYTOPLASMIC FACTORS AND INFECTIOUS AGENTS

Some cytoplasmic characters have turned out to be transmissible by infection. One of these is a peculiar sensitivity to carbon dioxide in a strain of Drosophila. Genetic tests with marked chromosomes showed that this factor is not carried on any chromosome. Usually the sensitivity was carried by the egg though occasionally by the sperm. Since the sperm cell has a little cytoplasm, this is not impossible. Then it was found that a fly could be infected with the sensitivity by injecting into it body fluid from a sensitive fly. Thus the factor is infective, though not by ordinary contact.

Another example is a factor causing mammary cancer in the mouse. This character is always transmitted through the female and was at first thought to be a case of cytoplasmic inheritance. A series of foster-nursing experiments showed that mice with mothers from the high cancer strain developed much fewer mammary cancers if they were separated from their mother at birth and nursed by mice from a noncancerous strain. Conversely mice of low cancer strain had increased cancer frequency if nursed by females with the high cancer strain. The factor is therefore carried by the milk. The active agent can be passed through a bacterial filter and appears to be a virus.

Another cytoplasmic agent has been found in a number of species of Drosophila. This is inherited with the egg cytoplasm and causes lethality in male embryos. Thus a female carrying the particle has almost exclusively female progeny. The occasional males are those that for some reason do not receive the particle from the mother; they are totally normal and never transmit the trait. The agent in this case is a spirochaete.

PROBLEMS

1. How could you test the idea that the carbon dioxide sensitivity factor in Drosophila might be carried by a supernumerary sperm (one that gets into the cytoplasm but doesn't reach the nucleus)?
2. In Paramecium, strain A is killer; strains B and C are sensitive. All three are pure strains and may be assumed to be homozygous. Strain A is mated with strain B; the progeny are allowed to multiply and several lines are derived, each from an animal that had undergone autogamy. All such lines were killers. When A was mated with C and the same procedure was followed, half the lines were killer and half sensitive. Designate for each strain its nuclear genotype and whether kappa was present in the cytoplasm.
3. An unknown strain D, when crossed with B gave after autogamy a ratio of three killer lines to one sensitive line; when crossed with C, a ratio of one killer line to three sensitive lines was obtained. What is the phenotype, genotype and cytoplasmic constitution of D?

   Note: Whether there is an exchange of cytoplasm depends on the length of conjugation. Assume, for the two problems above, that conjugation lasts long enough to permit cytoplasmic exchange.
4. Mitochondrial DNA has the property that both strands are transcribed. What would you expect if a mutation occurs?

REGULATION OF GENE ACTION

18

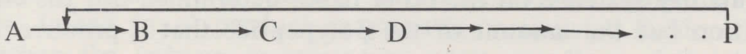
As the preceding chapters have shown, we know a great deal about how the genes are able to specify a particular sequence of amino acids in a polypeptide and thereby determine the specific kind of three-dimensional protein structure that is produced. Now the question arises: What regulates the amount of proteins and the times at which they are produced?

In higher organisms there is very little detailed genetic information about regulation of gene action. There is no doubt that genes are regulated and function in specific tissues at specific times. The orderly process of embryonic development and differentiation tells us that this must be so. At a more detailed level there are certain regions of polytene chromosomes in flies that produce conspicuous enlargements or "puffs." Some of these are RNA and, significantly, a specific chromosome region enlarges at a specific time in a specific tissue. Thus certain genes must somehow be stimulated to produce RNA at fixed times in fixed tissues, but the mechanism is not known.

Much of the current information comes from the study of enzyme systems, especially in microorganisms.

FEEDBACK INHIBITION, OR END-PRODUCT INHIBITION

For a number of biosynthetic pathways it is known that the accumulation of the final product of the reaction is capable of inhibiting the first step in the reaction, as shown in the diagram below.



The product, P, inhibits the enzyme that converts A into B. That this cannot simply be mass action is shown by the failure of intermediates to accumulate. In some cases the end product is known to act by altering the enzyme, probably in its three-dimensional structure, so that it is temporarily nonfunctional.

This is a very useful mechanism for it prevents overproduction of the product. When the product is scarce the synthesis takes place; when enough is produced, the synthesis stops.

However, the organism still carries all the enzyme proteins whether they are being used or not. It might be still more efficient to control the actual synthesis of the enzyme, so that the cell doesn't have to carry around any excess baggage in the form of unneeded proteins.

REGULATION OF ENZYME PRODUCTION

There are great variations in the amounts of enzymes and other proteins in different cells and at different times, despite the fact that the same genes are present. So there must be some means of regulating the amounts produced.

First, it is convenient to classify enzymes with regard to the way they are regulated. Enzymes may be:

1. Constitutive. Produced in fixed amounts, regardless of other substances in the cell.
2. Adaptive. Produced in varying amounts, depending on cell conditions.
    a. Inducible. Produced in response to an inducer, which is ordinarily the substrate on which the enzyme acts.
    b. Repressible. Repressed by the co-repressor, ordinarily the end product of the reaction.

For example, the enzyme beta galactosidase in *E. coli* is induced by its normal substrate, lactose. This is very useful to the bacterium because it produces the enzyme when it needs it—i.e., when it is in an environment containing lactose—and not otherwise. (The organism can be tricked into producing the enzyme by putting in compounds that are structurally similar to lactose, such as thiogalactoside. Some of these compounds induce the enzyme but are not hydrolyzed by it; this is called **gratuitous enzyme formation**. Such inducers have been very useful in the experimental analysis of the system.)

An example of the second type, control by repression, is the series of 10 enzymes responsible for histidine synthesis. When the product histidine is in low concentration all the enzymes are synthesized, but as the product increases in amount the enzyme synthesis stops.

STRUCTURAL AND REGULATORY GENES

We are already familiar with genes whose function is to specify a particular amino acid sequence, that is to determine the structure of a protein. Such genes are known as **structural** genes. **Regulatory** systems, on the other hand, determined not the composition but the amount of the polypeptide that is produced. A regulator gene may control several structural genes. In the histidine case the ten closely linked genes are all under the control of a common regulator. In other cases the genes may not be linked to the regulator, nor to each other.

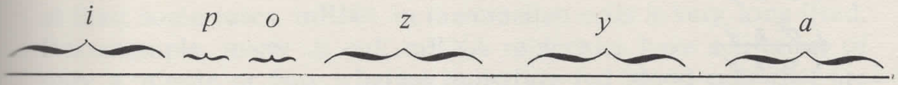
In general the mode of action of regulator genes is not understood. This is particularly true in higher organisms. But there is one system in *E. coli* that has been particularly well worked out and has become something of a paradigm for gene action studies. This is the operon concept developed by Jacob and Monod.

THE OPERON

The now classical example is the lactose locus in *E. coli*. By linkage analysis, it has been determined that there are six adjacent regions of the chromosome with related functions:

i^+ is the normal regulator gene
p^+ is the normal promoter
o^+ is the normal operator
z^+ is the normal structural gene for Beta-galactosidase
y^+ is the normal structural gene for permease, an enzyme that aids in getting lactose into the cell
a^+ is the normal structural gene for acetylase, whose function in this system is not known

The location of these regions is shown in the diagram below:



The whole series of structural genes governed by one operator (*o*) gene is called the **operon**.

A number of mutants of z^+ have been found at many mutant sites within the *z* region and all these change the specificity of the galactosidase. Most of them render it nonfunctional so that the active enzyme is not produced at all. These mutants will be designated z^-. Similar mutants are found for the permease locus, *y*, and for acetylase, *a*.

The *i* and *o* regions were discovered by finding mutants that change their normal function. When i^+ mutates to i^- or to a deletion the genes z^+, y^+, and a^+ are no longer regulated, and produce enzyme in full amounts whether the inducer is present or not (that is to say, constitutively). The regulator gene, i^+, produces a protein that shuts off the activity of the y^+, z^+, and a^+ genes. However, an inducer such as lactose inactivates the repressor protein so that it is no longer effective.

The operator gene (o^+) affects only those loci that are linked to it in the cis phase. The repressor acts through the operator. This is shown by a mutant; when o^+ is replaced by the mutant o^c the genes linked to o^c produce enzymes in unregulated fashion (i.e., whether the inducer is present or not) regardless of the state of the *i* locus. The o^c stands for constitutive, since the enzymes are produced constitutively when it is present.

I shall illustrate these principles, as well as show the evidence on which this hypothesis is based, with a table showing the phenotypes corresponding to some of the genotypes. As you know, *E. coli* is normally haploid so it is not possible to study dominance relations. However, in this case one can get diploidy for the appropriate region since one F′, (an F factor carrying an extra bit of bacterial chromosome mentioned in Chapter 9), carries this region of the chromosomes.

In the table on the following page, + means that the enzyme is produced; - means that it is not in significant amounts.

	Inducer present		*Inducer absent*	
Genotype	*Galactosidase*	*Permease*	*Galactosidase*	*Permease*
Haploids				
$i^+ o^+ z^+ y^+$	+	+	−	−
$i^+ o^+ z^- y^+$	−	+	−	−
$i^+ o^+ z^+ y^-$	+	−	−	−
$i^- o^+ z^+ y^+$	+	+	+	+
$i^+ o^c z^+ y^+$	+	+	+	+
$i^+ o^c z^+ y^-$	+	−	+	−
Diploids				
$\frac{i^+ o^c z^+ y^-}{i^- o^+ z^- y^+}$	+	+	+	−
$\frac{i^- o^c z^+ y^-}{i^+ o^+ z^- y^+}$	+	+	+	−
$\frac{i^+ o^+ z^+ y^-}{i^- o^c z^- y^+}$	+	+	−	+

From these data one can see that, in general for structural genes, + is dominant to −. This is as expected; half as much enzyme still works. Also i^+ is dominant to i^- as expected. From the last three genotypes it is clear that the cis-trans phase makes no difference as far as the *i* locus is concerned, but that the *o* locus governs only those genes that are on the same strand with it.

We can now discuss this system in molecular terms.

The promoter is the point to which the RNA polymerase attaches to begin the process of transcription. The polymerase molecule can be separated into two units, the **sigma** factor which is thought to recognize the proper DNA site, and the rest, the **core enzyme**, that seems to carry out the process of synthesis. Mutants that map at the promoter site lead to less efficient transcription.

The regulator (i^+) gene produces the repressor. This is a protein, synthesized in the usual way, which is specifically able to bind to the operator site. There it somehow blocks the process of transcription so that mRNA is not produced. The inducer, in this case lactose or a lactose derivative, attaches to the repressor and makes it incapable of functioning. Thus, in the presence of the inducer, transcription takes place. The o^c mutant changes the operator region so that the repressor no longer attaches.

The synthesis of mRNA starts at the promoter (although the promoter itself is not transcribed) and continues through all three structural genes. Thus a single messenger is produced for all three genes, each of which has its own start and stop codons for protein synthesis. Such polycistronic messengers are produced for other groups of genes under control of a single operator. The histidine operon includes 10 genes, all transcribed onto a single messenger.

This is a very efficient system, for it means that a group of genes with related functions can be turned on or off together. In an inducible system, the inducer starts the transcription of all of them; in a repressible system the co-repressor shuts the process down.

Although there are a number of instances in microorganisms where there are closely linked genes governing a sequence of related chemical steps, this is not typical of eukaryotes. Genes with related functions are scattered among the various chromosomes.

Presumably every mRNA is associated with a promoter, for there has to be some site for the transcribing enzyme to recognize. However, not all loci in *E. coli* have an operator. Whether gene regulation in higher organisms is done by something like an operator, or whether the systems are totally different remains to be determined.

Bacterial mRNA is often quite unstable. On the other hand, in at least some cases, mRNA in mammalian cells is very long lived. For example, many *E. coli* mRNA molecules have a lifetime of only a minute or two, whereas immature red blood cells that are producing hemoglobin have messenger RNA's with a very long life. This makes sense, for the environment of a red blood cell is very constant compared with that of *E. coli* which often finds itself suddenly changed from a child's intestinal tract rich in lactose to the outside world with entirely different nutrients.

NONTRANSLATED DNA

It is now well known that only a very small fraction of the DNA codes for proteins in the higher eukaryotes. In Drosophila, as we saw in Chapter 15, there is some 30 times as much DNA as is translated. In the mammals the ratio is probably very much higher than 30, but uncertainty about the gene number prevents an exact statement.

It is also known that a great deal more DNA is transcribed than is translated. What all this excess DNA and RNA does is still largely a mystery. Some of the DNA is highly redundant; the same region may be repeated hundreds of thousands of times. But such redundant DNA makes up less than 1/4 of the total; there is still a large excess of unique-sequence DNA.

It is an obvious thought that much of this is used for regulatory functions, for the excess is correlated with the complexity of the organism. To coordinate all the diverse activities in an enormous number of cells during embryonic development and later in life must require very complicated mechanisms. But whatever they are, they are still largely unknown.

Different species are often very similar as regards their proteins. For example, man and the chimpanzee are alike for more than 99% of the amino acid sites in corresponding proteins. The similarity in exact amino acid composition gets less as the species are less related, but homologous proteins still remain very similar in their functional activity.

All this suggests that the major differences among the different vertebrates have more to do with regulating the amounts of the different proteins in different cells than they have to do with producing different proteins. The differences between humans and chimpanzees would seem to be mainly regulatory. Supporting evidence comes from the fact that the total DNA seems to differ more in related species than do the individual proteins. On the other hand, this could also be understood by assuming that less precise nucleotide sequences are needed for regulation than for amino acid specification, and so random differences are likely to occur more often in regulatory parts of the DNA.

PROBLEMS

1. For the following genotypes, tell whether galactosidase and permease are produced with and without an inducer:

   a. $i^- o^+ z^+ y^- / i^- o^c z^- y^+$
   b. $i^+ o^+ z^- y^- / i^- o^c z^+ y^+$

2. Frequently the genes produce more enzyme when they are on the F′ particle than when they are on the chromosome. Suggest an explanation.

3. There is another mutation, that by linkage maps is very close to o^c. This prevents all enzyme production by genes linked to it in the cis phase, whether or not the inducer is present. Suggest a hypothesis.

4. There are three linked loci in *E. coli*, their order having been determined by linkage tests. In the diagram below the mutants are given in correct order; *m* indicates the mutant and + the normal. P means enzyme present in large amounts; A means it is absent or present only in a trace. One of the mutants is a regulator, one an operator, and one is structural. Which is which?

	Phenotype	
Genotype	Inducer present	Inducer not present
m + +	P	P
+ *m* +	A	A
+ + *m*	P	P
+ + +/*m m m*	P	A
+ + *m*/*m m* +	P	P
+ *m* +/*m* + *m*	P	P
m + +/+ *m m*	P	A

IMMUNOGENETICS

19

ANTIGEN-ANTIBODY REACTIONS

When a foreign substance called an **antigen** is brought into the body it stimulates the appearance of **antibodies**. Antibodies are a type of protein, gamma globulin. Antibodies will react with the antigen whenever it is again introduced in the body, or when antigen and antibody are mixed in a test tube.

There are various kinds of detectable reactions between antigen and antibody, but they are all thought to be secondary manifestations of the same basic phenomenon. If the antigen is part of a cell, the reaction may cause the cells to clump (agglutination), or be destroyed (lysis). Sometimes the reaction is detected by a precipitation. Also the rejection of a skin or organ transplant is due to an antibody reaction.

The general principles of antigen inheritance are as follows:

1. Every antigen is due to a dominant gene, i.e., the antigen is present whether the gene is homozygous or heterozygous. Heterozygotes produce both antigens; this is sometimes called co-dominance.
2. Normally, an individual possesses or can produce antibodies only against those antigens that he does not have.

The classical genetic studies of antigen-antibody reactions involve the blood groups. These are among the best known human polymorphisms, in addition to their importance in transfusions and some kinds of diseases. So we begin our discussion of immunogenetics with the classical blood groups.

THE A-B-O SERIES

In 1900 Landsteiner observed that when blood cells from one person were mixed with serum from another a clumping or **agglutination** of the cells sometimes occurred. He was soon able to divide the human population into four groups on the basis of these agglutination tests. The antigens were found to be on the cells and the antibodies in the blood serum. A person possesses antibodies for those antigens which he does not have. The basic facts are summarized in the table on the following page.

To determine the group to which a person belongs a sample of his blood cells is mixed with serum containing anti-A antibodies and another sample is mixed with anti-B serum. (The anti-A serum could be obtained from a group B person, the anti-B serum from a group A person.) If the unknown blood is clumped by both kinds

Table 19-1

Group	Genotype	Reaction with antibodies		Blood contains		Frequency in U.S. white population
		anti-A	anti-B	Cellular antigens	Serum antibodies	
O	*O/O*	−	−	none	anti-A anti-B	45%
A	*A/A, O/A*	+	−	A	anti-B	42%
B	*B/B, O/B*	−	+	B	anti-A	10%
AB	*A/B*	+	+	A, B	none	4%

of sera it is AB; if by neither, it is O; if by anti-A, it is A; if by anti-B, it is B. This can be seen from the third and fourth columns of the table above.

When transfusions of incompatible blood are given serious reactions occur. The incoming cells are agglutinated by the serum of the recipient causing capillary blockage and cell destruction. This can be prevented by ascertaining that cells of the donor contain no antigen for which the recipient has antibodies in his serum.

The manner of inheritance of the characters can be determined directly from the table of genotypes above. As can be seen there is a single genotype, *O/O*, corresponding to group O and a single genotype, *A/B*, for group AB. However, the groups A and B can be either homozygous or heterozygous. Predicting the blood groups of children from that of their parents is possible within certain limits. For example, two parents of group O could have only group O children. Children whose mother was O and whose father was A could not be B or AB. On the other hand, two parents, one of group A and the other of group B, could produce children of any group. This kind of knowledge has practical application in disputed paternity and other medico-legal cases.

A great deal is now known about the chemistry of the blood cell antigens. They are glycoproteins and glycolipids associated with the red blood and other cell surfaces. The basic molecule is the H antigen and the final step in its synthesis is governed by the *H* gene, which is at another locus independent of the ABO alleles. This antigen was not discovered until fairly recently because there is no normal antibody against it. The reason for this, in turn, is that almost everybody has the antigen; so antibodies are not produced. Persons that are homozygous recessive, *hh*, fail to produce the antigen, nor do they produce A or B antigens, even if they have the *A* and *B* genes. This is called the Bombay phenotype because it was first found in that part of India. It is very rare, even there, with an incidence of less than 1/10,000, but is even rarer in other parts of the world. The H antigen (or some of it) is converted into A and B antigens under the influence of the *A* and

B loci. The *A* gene is responsible for the enzyme that adds a galactosamine to the antigen molecule while the *B* gene is responsible for the enzyme that adds galactose. Whether these are structural genes or some kind of regulatory gene is not known.

There are now more than 15 well established blood group systems. I shall mention only two of them, the first (the M-N series) because of its early discovery and wide use in the study of human populations, and the second (the Rhesus factors) because of its association with hemolytic disease. Neither of these is understood in the chemical detail of the ABO system.

THE M-N SERIES

In 1928 Landsteiner and Levine, by injecting human blood into rabbits, produced antibodies by which it was possible to further subdivide the human population. These new antigens were inherited independently of the ABO series and were named M and N. Antibodies against these antigens are not ordinarily found in human blood; therefore, these antigens are of no importance in transfusions, but are of interest as genetic phenomena.

Table 19-2

Group	Genotype	Reaction with antibodies		Blood contains		Frequency in U.S. white population
		anti-M	anti-N	Cellular antigens	Serum antibodies	
M	*M/M*	+	−	M	none	26%
MN	*M/N*	+	+	M, N	none	53%
N	*N/N*	−	+	N	none	21%

THE RHESUS (RH AND HR) FACTORS

In 1940 Landsteiner and Wiener reported the discovery of yet another set of blood antigens. These were found by a still different technique. They injected blood from a rhesus monkey into rabbits, and after the antibodies for previously known human antigens had been removed, the rabbit serum still agglutinated the cells of about 85% of the humans whose blood was tested. Thus these 85% contained a blood antigen in common with the rhesus monkey and the factor was named Rh.

It was soon discovered that this factor was inherited as a simple Mendelian dominant. The gene has been designated by two different systems. Wiener called it Rh_O while Fisher in England proposed a different nomenclature in which this antigen was called D. I shall use the Fisher system.

Previous to the discovery of the Rh factors it had been noted that there were several cases of transfusion reactions which could not be accounted for on the basis of the A and B antigens. These reactions almost always occurred in persons who had had repeated transfusions or in women who had had several pregnancies. In many of the latter cases one or more pregnancies had terminated in stillbirth or abortion. Wiener and Levine associated all these phenomena with the Rh system. This has since been confirmed.

Table 19-3

Group	Genotype	Reaction with anti-D antibody	Blood contains: Cellular Antigens	Blood contains: Serum Antibodies	Frequency in U.S. white population
Rh$^+$	*D/D, D/d*	+	D	none	85%
Rh$^-$	*d/d*	–	none	none*	15%

*Does not normally contain antibodies but is capable of producing anti-D antibodies on stimulation with cells containing the D antigen.

Rh negative individuals do not normally possess antibodies to Rh positive cells but are capable of producing such anti-D antibodies when the D antigens are repeatedly introduced. This happens in two ways. The first is through repeated transfusion. If an Rh negative person is given several transfusions of Rh positive blood he may eventually build up antibodies which react with the cells at some future transfusion with Rh positive blood. The second way in which antibodies are formed is in the case of an Rh negative woman carrying an Rh positive embryo. The D antigens of the fetus somehow stimulate the mother to produce anti-D antibodies. These antibodies circulate through the placenta and destroy many of the fetal blood cells, causing a hemolytic disease, known as erythroblastosis foetalis. Fortunately, the mother becomes immunized only rarely, so the disease is much less common than would be expected from the frequency of Rh$^-$ women with Rh$^+$ children.

Further studies have greatly increased our knowledge of the Rh system and it is much more complex than previously thought. Wiener, one of the discoverers of the Rh factor, suggests multiple alleles. An alternative hypothesis, suggested by Fisher, is that there are three sets of linked genes. Since no clear cases of crossing-over have occurred it is not possible to demonstrate separate loci, but we shall use this notation because of its greater simplicity. The following table of equivalents will aid in keeping the terminology straight if you read references by different authors, some of whom use one system and some the other.

ANTISERA: Wiener	ANTISERA: Fisher	GENETIC UNITS: Wiener	GENETIC UNITS: Fisher
anti-rh$_o$	anti-D	R^z	*CDE*
anti-rh′	anti-C	R^1	*CDe*
anti-rh″	anti-E	R^2	*cDE*
anti-hr$_o$	anti-d		
	(hypothetical)	R^o	*cDe*
anti-hr′	anti-c	r^y	*CdE*
anti-hr″	anti-e	r''	*cdE*
		r'	*Cde*
		r	*cde*

These genes (or sets of genes) occur in any person in diploid pairs, so that one might have any combination of two of the units. This makes a total of 36 genotypes of which 27 are serologically distinguishable. Some of the most common genotypes and the kinds of antisera to which they react are given in the table below. The general principle is that cells of any genotype will react with all the antisera for which it contains genes, whether the genes are present in single or double dose.

Table 19-4

	Genotypes	Freq.	Reaction with anti-C anti-Rh′	anti-D anti-Rh$_o$	anti-E anti-Rh″	anti-c anti-Hr′	anti-e anti-Hr″
Rh$^-$	*cde / cde*	14.36	–	–	–	+	+
	cdE/cde	1.23	–	–	+	+	+
Rh$^+$	*cDe/cde*	2.31	–	+	–	+	+
	CDe/CDe	19.02	+	+	–	–	+
	CDe/cDe	2.66	+	+	–	+	+
	CDe/cde	33.06	+	+	–	+	+
	cDE/cde	9.70	–	+	+	+	+
	CDe/cDE	11.16	+	+	+	+	+

A particular antibody can be made only by a person not possessing the corresponding antigen. For example, anti-D can be made only by a *dd* person who has been immunized with cells containing the D antigen (*DD* or *Dd* cells). Over 90% of the cases of erythroblastosis are due to the anti-D factor. Most of the rest are due to anti-C.

Hemolytic disease due to Rh incompatibilities is rapidly becoming preventable. Its incidence can be reduced by following the now standard practice of not giving Rh$^+$ blood to Rh$^-$ women. Furthermore it is possible to greatly decrease, if not eliminate sensitization of the Rh$^-$ mother by introducing incomplete antibodies against Rh$^+$ blood cells immediately after she has given birth. This blocks any fetal red blood cells that may have gotten into her blood stream so that they cannot act as antibody stimulators.

There is an interesting interrelationship between the ABO and the Rh groups, for ABO incompatibility seems to protect from Rh-induced hemolytic disease. The following data are from Levine (*Human Biology* 30:14-28).

Table 19-5
Number of cases of Rh hemolytic disease

ABO Incompatible			ABO Compatible		
O mother x	A father	319	A mother x	O father	765
O	B	81	B	O	156
A or B	AB	33	AB	A or B	58
O	AB	0	AB	O	59

The last row is the most significant, for in the O mother and AB father every embryo has an antigen for which the mother has a corresponding antibody. It thus appears that when the child and mother are incompatible for ABO factors, the child is protected from erythroblastosis due to the Rh factor. Perhaps in this case, any fetal cells that leak through the placenta are destroyed by the anti-A or anti-B antibodies before they can persist long enough to stimulate the formation of anti-Rh antibodies.

BLOOD GROUP MOSAICS IN CATTLE

Cattle blood groups have been extensively studied. There are many systems and many known types within each system and they follow the same rules of co-dominant inheritance as the human blood groups. The number is now large enough that each individual is almost unique; it is usually possible to identify an individual from a sample of its blood. This knowledge is widely used to establish parentage in doubtful cases.

It has long been known that when opposite-sexed twin calves are born, the female develops into a sterile intersex-like individual known as a **freemartin**. The explanation became clear when it was discovered that in such cases each twin has a mixture of both kinds of blood, as identified by the blood groups. During embryology the circulatory systems of the twins are connected so that various kinds of cells from the two individuals get mixed. Somehow, having a large number of cells from her brother suppresses the sexual development of the female.

The most striking thing about these twins with mixed blood cells is that there are no incompatibilities from the different blood groups. One might expect frequent effects comparable to those of transfusion mismatches. This led R. D. Owen to realize that when antigens are present early enough in embryology the antibodies are not stimulated. This phenomenon of **immune tolerance** helps explain why an individual does not have antibodies to his own antigens. If an embryo is sufficiently exposed to an antigen, he is forever lacking antibodies to this antigen, even with repeated stimulation later.

TISSUE TRANSPLANTS

A long series of experiments with mice and guinea pigs have shown that the ability to accept or reject tissue transplants is a genetic property. Transplants of skin or other tissue can be made readily from one part of the body to another, and from one identical twin to another, but not between unrelated individuals. An exception to this is found in highly inbred mice, where all the individuals are virtually identical genetically.

The rules of transplant acceptance are:

Donor	Recipient	Accept or reject
Inbred strain	Same strain	Accept
Inbred strain	Another strain	Reject
Inbred strain	F_1 hybrid between this and another inbred strain	Accept

F_1 hybrid	Inbred strain	Reject
F_2 hybrid	F_1 hybrid	Accept
F_1 hybrid	F_2 hybrid	Reject (except in small minority)

These results can all be interpreted simply with the following rules: There are a large number of co-dominant histocompatibility genes. A recipient will accept a graft only if the donor tissue has no genes not possessed by the recipient.

For example, if inbred strain 1 is of genotype *aa BB CC dd EE* and strain 2 is *AA bb cc DD ee*, the F_1 hybrid is *Aa Bb Cc Dd Ee*. The F_1 hybrid has all the factors that either parent has, so will accept grafts from either. However, since the F_1 has factors not possessed by the pure strains, it cannot donate to them.

It is possible to estimate how many histocompatibility loci there are by the proportion of acceptances when grafts are made from F_1 to F_2. There are now known to be at least 35 loci in the mouse. However, one locus, called *H-2*, is much more important than any others; in fact it is more important than the others put together when grafts are made along with immunosuppressive drugs. If mice are matched for the *H-2* locus, the rejection can be greatly postponed or even prevented by drugs, despite incompatibilities at several other loci.

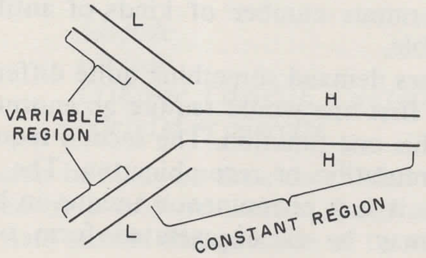
The situation seems to be very similar in humans. There is one major region, much more important than the others. There are two major loci in this region, which are about 1 map unit apart on the chromosome. By matching for these two loci, the survival of transplants is greatly enhanced.

There are a very large number of alleles at each of the loci, so most persons will be heterozygous at both loci. This means that sibs will be identical for this locus 1/4 of the time (ignoring the rare crossovers). By choosing such sibs and using immunosuppressive drugs, kidney graft survival rates are becoming more and more successful.

THE PROBLEM OF ANTIBODY VARIABILITY

Antibodies are remarkable entities. The organism seems to have potentially available an enormous number of kinds of antibodies, enough to react with almost any kind of foreign substance. So far we have concentrated on the genetics of antigens, but antibodies are also under genetic control.

Antibodies are **immunoglobulins** and all have a characteristic molecular structure.



There are two kinds of polypeptides, one short or light (L) and one long or heavy (H). These join together, held by disulfide bonds from sulfur-containing amino acids, into the Y-shaped structure shown above.

Some of the regions of the molecule are variable, such that there are many amino acid differences in different molecules, even in cells from the same person. The other regions are relatively constant, although there is some variability from person to person. The variable region is what makes antibodies capable of responding to so many kinds of antigens. It is known from electron microscope studies that the antigen attaches to the opening of the V (at the left end of the diagram), so it is not surprising that the variable part of the molecule is at this end. The different amino acid sequences cause a variety of three-dimensional structures which permit attachment to antigens with complementary structures.

There are at least five general classes of immunoglobulins, each controlled by different, but presumably very similar, genes. But any particular cell has only one type. Even in heterozygotes, only one of the alleles is expressed. So each cell produces a particular immunoglobulin. The best evidence for this comes from the study of some kinds of malignant disease that cause one particular cell to multiply enormously so that it becomes the prevailing type. In such cases, the malignant cells from a single person are all identical and have a single immunoglobulin; but each person with the disease produces a different type of immunoglobulin.

A normal person, then, has many kinds of antibody carrying cells. When an antigen enters the body, one particular type of cell—the one with an immunoglobulin that happens to be shaped to fit the antigen—is stimulated to multiply. Thus, the person becomes immune to this antigen.

The mystery is how all this variability is generated. It is clear that some of the differences are inherited, for even the more constant parts of the molecule differ from individual to individual. There are two alternative hypotheses: (1) this assumes that there are an enormous number of similar genes, each capable of producing its own type of immunoglobulin. Then, during development, all but those producing a single type of immunoglobulin are suppressed, so each cell produces only a single type; (1) the alternative assumes that only a small number of genes are involved, but that, by somatic mutation and possibly recombination, enough variability is introduced in embryonic cell divisions to generate the enormous number of kinds of antibodies that are potentially available.

Both hypotheses demand something quite different from most of genetics. The first one would require an enormous number of genes with just this one function. The second requires an unusual rate of somatic mutation or recombination. The question is not resolved; perhaps it is a phenomenon unique in biology. On the other hand, it may be an exaggerated form of some regular

---

phenomenon
differentiatio
enormously c

1. An Rh^-
   children.

2. Which of

   Child
   a. O, M
   b. O
   c. A, M
   d. AB,
   e. B, N,

3. To what
   From wh
   sion?

4. An Rh^-
   and has a
   is a pro
   What kin
   zygous o

5. What pro

6. What gen

7. The follo

   Woman
   Child
   Man 1
   Man 2

   Which m

8. In anoth
   tests were

   Mother
   Child
   Man 1
   Man 2

   Can a de
   two chil
   additiona

phenomenon and may offer clues to other types of embryonic differentiation by which the descendants of a single cell have such enormously diverse functions.

PROBLEMS

1. An Rh$^-$ woman with a heterozygous Rh$^+$ husband has three children. What is the probability that all three are Rh$^+$?

2. Which of the following are possible, and which are not?

	Child	Mother	Father
a.	O, M	A, M	B, MN
b.	O	O	AB
c.	A, MN, Rh$^+$	O, N, Rh$^-$	AB, M, Rh$^-$
d.	AB, M, Rh$^-$	AB, N, Rh$^+$	A, MN, Rh$^+$
e.	B, N, Rh$^+$	O, MN, Rh$^-$	O, N, Rh$^+$

3. To what groups can a person of group A give a transfusion? From what groups can a person of group A receive a transfusion?

4. An Rh$^-$ woman has given birth to an erythroblastotic infant and has anti-D antibodies. If the husband is heterozygous there is a probability of 50% that the next child will be normal. What kind of serum would tell whether the husband is homozygous or heterozygous for the factor D?

5. What proportion of the U.S. white population are O, Rh$^-$?

6. What genotypes could react with all six Rh antisera?

7. The following data are from a recent court case.

Woman	A, M, *cde*/*cde*
Child	A, MN, *cde*/*cde*
Man 1	O, M, *CDe*/*CDe*
Man 2	A, N, *cdE*/*cde*

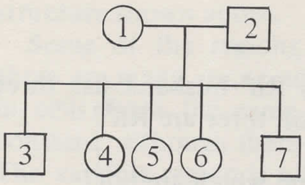
Which man could be the father?

8. In another case the Rh data are as follows (the AB, MN, and P tests were inconclusive).

Mother	*cde*/*cde*
Child	*cde*/*cde*
Man 1	*CDE*/*cde*, *Cde*/*cDE*, *CdE*/*cDe*, or *CDe*/*cdE*
Man 2	*cDe*/*cDe* or *cde*/*cDe*

Can a decision be made on these data? Man number 1 also had two children by a *cde*/*cde* woman, both *cDE*/*cde*. Do these additional data answer the question as to which is the father?

9. Fill in the genotypes of the following pedigree.



Serological data

	anti-C	anti-D	anti-E
1.	+	+	–
2.	+	–	+
3.	–	–	–
4.	+	+	+
5.	–	–	–
6.	+	–	+
7.	+	+	–

10. If an anti-D serum becomes available, predict how the following cells will react with it: *cde/cde, cdE/Cde, CDe/cde, CDe/CDe.*

11. Assume that you are in charge of an outrageously careless lying-in hospital. Four children are born with the following blood groups, but the nurses don't know which parents they go with. Assign them to the proper families.

		Parents	Father	Mother
Baby A	O, MN, Rh$^-$	1	B, N; Rh$^-$	A, MN, Rh$^-$
Baby B	AB, N, Rh$^+$	2	B, N, Rh$^+$	AB, N, Rh$^-$
Baby C	B, MN, Rh$^+$	3	AB, MN, Rh$^-$	O, MN, Rh$^+$
Baby D	O, N, Rh$^-$	4	A, M, Rh$^+$	A, N, Rh$^+$

12. Suppose you cross two inbred strains of mice which differ from each other with respect to five pairs of histocompatibility genes. What proportion of transplants would be accepted in each of the following:

	Donor	Recipient
a.	inbred strain 1	inbred strain 2
b.	inbred strain 2	F_1 hybrid
c.	F_1 hybrid	inbred strain 1
d.	F_2	F_1
e.	F_1	F_2
f.	F_2	inbred strain 1
g.	F_2	F_2
h.	inbred strain 2	F_2

13. Suppose there were only one histocompatibility locus, but that this had a very large number of alleles. What would be the expected proportion of acceptances of grafts between unrelated individuals? parent and offspring? sibs? (And for fun: double first cousins? children of identical twins? children of identical twins who marry identical twins?)

l data

anti-E

−
+
−
+
−
+
−

ıow the follow-
CDe/cde, CDe/

eously careless
the following
h parents they

Mother
A, MN, Rh⁻
AB, N, Rh⁻
O, MN, Rh⁺
A, N, Rh⁺

which differ
stocompatibil-
be accepted in

ent
rain 2
l
rain 1

rain 1

y locus, but
would be the
etween unre-
And for fun:
? children of

POPULATION GENETICS

20

RANDOMLY MATING POPULATIONS

In some cases, such as in the study of animals or plants in nature or the human population, experimental matings are not practical and one must resort to other methods. For such studies special procedures have been devised. If the population can be assumed to be mating at random the procedures are greatly simplified, and fortunately this is often the case.

When we say a population is mating at random with respect to a certain character we mean that an individual possessing that trait is no more nor less likely to mate with another with the trait than would be determined by chance, given the frequency of the trait in the population. For example, most human populations can be considered to be mating at random with respect to blood group genes.

On the other hand, the human population is not mating at random with respect to a number of factors, such as the skin color genes. Such a departure from randomness in the direction of similarity of mates is called **assortative** mating. An example of assortative mating occurs with height in humans. There is a tendency for tall to marry tall, and short to marry short. There is also a strong tendency for assortative marriage in intelligence; the similarity between husband and wife in test scores is as great as that between parent and child. There is also the theoretical possibility of disassortative mating, but examples are rather rare. The folk lore that opposites attract doesn't seem to be true in either animal or human populations (except for the character sex itself!).

Notice that it is possible for a population to be mating at random with respect to one character and assortatively for another at the same time.

Another departure from random mating is due to **inbreeding**. By inbreeding we mean mating between individuals more closely related than a pair randomly chosen from the population. There is a slight amount of inbreeding in the human population due to the fact that marriage of cousins and second cousins is more frequent than would occur if mating were at random. The opposite of inbreeding is called outbreeding. This occurs when mating is between individuals who are less closely related than a random sample of the population.

THE HARDY-WEINBERG PRINCIPLE

The fundamental idea of gene distribution in populations was developed independently by Hardy, a British mathematician, and Weinberg, a German physician, in the year 1908. It is a deduction from the Mendelian principles of heredity and the assumption of random mating. It may be stated as follows:

If in a large population, p_1 is the proportion of gene A_1, and p_2 is the proportion of gene A_2, then after one generation of random mating the genotypes will attain and remain at the following frequencies:

Genotype	Frequency
A_1A_1	p_1^2
A_1A_2	$2p_1p_2$
A_2A_2	p_2^2

This can easily be demonstrated by thinking of random mating as being equivalent to having each parent contribute equally to a pool of gametes, from which random pairs are drawn. The results are as shown below.

		Sperms			
		p_1	A_1	p_2	A_2
Eggs	p_1 A_1	p_1^2	A_1A_1	p_1p_2	A_1A_2
	p_2 A_2	p_1p_2	A_1A_2	p_2^2	A_2A_2

More generally, among a series of multiple alleles the frequency of any particular homozygote is the square of the frequency of the gene, and any particular heterozygote is twice the product of the frequencies of the two genes making the heterozygote. Formally,

Genotype	Frequency
A_iA_i	p_i^2
A_iA_j	$2p_ip_j \quad (i = j)$

where i and j stand for any desired numbers from 1 to the total number of alleles.

Example: About 1 child in 10,000 is born with phenylketonuria. This is known to be due to a recessive gene, *a*. What proportions of the normal population are carriers of this gene?

If we let p_2 be the frequency of the gene for phenylketonuria, then $p_2^2 = 1/10{,}000$, or p_2 is the square root of this, about 1/100. The frequency of the normal gene must be 1-1/100, or 99/100. The proportion of heterozygous carriers in the population must then be 2(1/100) (99/100), or 198/10,000, roughly 1 in 50 persons.

Notice that there are 198 times as many persons who are heterozygous carriers of the gene as there are who have the gene in the homozygous condition.

Another Example: Assume a population in which the blood group genes O, A, and B are in the proportions .6, .3, and .1. If marriages occur at random, what will be the frequencies of persons with the four blood groups?

Group O:	$.6^2$			= .36
Group AB:	2(.3)(.1)			= .06
Group A:	Homozygous, *AA*	$.3^2$	= .09	
	Heterozygous, AO	2(.3)(.6)	= .36	
	Total group A			= .45
Group B:	Homozygous, *BB*	$.1^2$	= .01	
	Heterozygous, *OB*	2(.6)(.1)	= .12	
	Total group B			= .13
				1.00

MORE THAN ONE LOCUS

If a population is mating at random and has reached equilibrium, the frequencies of composite genotypes are the products of the component single locus frequencies. For example,

If p_1 is the frequency of gene A_1
If p_2 is the frequency of gene A_2
If r_1 is the frequency of gene B_1
If r_2 is the frequency of gene B_2

then, the frequency of genotype $A_1A_1B_1B_1$ is $(p_1^2)(r_1^2) = p_1^2r_1^2$

$A_1A_1B_1B_2$ is $(p_1^2)(2r_1r_2) = 2p_1^2r_1r_2$

$A_1A_2B_1B_2$ is $(2p_1p_2)(2r_1r_2) = 4p_1p_2r_1r_2$

and so on.

It should be emphasized that, while the equilibrium frequencies for a single locus regardless of the number of alleles is reached in a single generation of random mating, the approach to equilibrium for more than one locus is only gradual. For unlinked genes the approach is fairly rapid, the departure from equilibrium proportions being reduced by half each generation; but for closely linked genes the disequilibrium may persist for many generations.

Hence, although it is reasonable to assume random mating proportions for a single locus, when multiple loci are concerned this cannot be assumed unless the population is known to have been mating randomly for some time.

Example: In the United States the frequency of the O blood group factor is .67 and that of the D factor in the Rh series is .60. What is the frequency of persons who are group O and Rh-positive, assuming a long history of random mating?

Frequency of group O is $.67^2$		= .45
Frequency of *DD* is $.60^2$	= .36	
Frequency of *Dd* is 2(.60)(.40)	= .48	
Total Rh-positive		= .84
Frequency of O, Rh+ .45 x .84	= .378 or about 38 percent.	

GENE FREQUENCY ANALYSIS

The determination of the manner of inheritance of traits in the human population is often very difficult. There are two procedures that are frequently followed: (1) the examination of enough pedigrees so that the desired matings are found among them, and (2) analysis of gene frequencies on the assumption that the population is at equilibrium according to the Hardy-Weinberg rule.

An example of the use of gene frequency methods for testing a genetic hypothesis is the study of the inheritance of the M and N blood groups. Out of 420 persons tested 137 were M, 196 were MN, and 87 were N. One hypothesis (the correct one, as you know) is that the group M is homozygous *M/M*, MN is heterozygous *M/N*, and N is homozygous *N/N*. Since each *M/M* person has two *M* genes and each *M/N* person has one per cell, the proportion of *M* genes in the population is $\frac{2(137) + 196}{2(420)} = .56$; $1 - .56 = .44$. The expected proportions of the three genotypes are $(.56)^2$ *M/M*, $2(.56)(.44)$ *M/N*, and $(.44)^2$ *N/N*.

The actual and expected numbers are tabulated below:

Table 20-1

Genotype	Expected proportion	Expected number	Observed number
M/M	.314	132	137
M/N	.493	207	196
N/N	.194	81	87
		420	420

$\chi^2 = 1.22$
D. F. = 1
Prob. = .28

The large probability shows that these data are consistent with the hypothesis, whereas if we test the same data with any other simple hypothesis the difference between observation and expectation is highly significant.

Similar but more complicated analyses can be used to test other hypotheses of the nature of inheritance of human traits. In general the procedure is to formulate the hypothesis, compute the gene frequencies on this assumption, and compare the expected proportion of offspring from certain crosses with those actually observed.

MULTIPLE ALLELES

The principle is unchanged. Here are some data from Selander (Univ. Texas Pub. 7103) on the frequency of esterase enzyme types. There are three alleles, designated *a, b,* and *c,* in this population of field mice.

We compute the allele frequencies as before by counting each homozygote twice and each heterozygote once, and dividing by the total number of genes. Thus $p_a = (4 + 13 + 3)/194 = .103$. Then we compute the expected numbers from the Hardy-Weinberg

ce of traits in the
ere are two pro-
examination of
re found among
assumption that
Hardy-Weinberg

ods for testing a
of the M and N
ere M, 196 were
ect one, as you
, MN is hetero-
ach M/M person
ne per cell, the
+ 196
0) = .56; 1 -
e genotypes are
below:

d
r

$\chi^2 = 1.22$
D. F. = 1
Prob. = .28

consistent with
vith any other
n and expecta-

d to test other
aits. In general
pute the gene
expected pro-
hose actually

rom Selander
erase enzyme
nd c, in this

ounting each
d dividing by
/194 = .103.
rdy-Weinberg

Table 20-2

	Genotypes and numbers						Total	Allele frequencies		
	aa	*bb*	*cc*	*ab*	*ac*	*bc*		P_a	P_b	P_c
Observed	2	53	5	13	3	21	97	.103	.722	.175
Expected	1.0	50.5	3.0	14.4	3.5	24.5	96.9			
	$\chi^2_3 = 3.0$, Prob. = .4									

principle. As can be seen, the data are in good agreement with the expectation for random mating.

X-LINKED GENES PROBLEMS

Since the female has two X chromosomes, the genotype proportions in a randomly mating population are the same as for an autosome. However, in the male where there is only one X chromosome the frequencies are simply those of the gametes. If you are dealing with an organism like birds or butterflies where the female is XY, the conclusions are of course reversed.

The most important consequence of this is for a rare recessive gene. If the frequency of affected males is 0.01, this means that the proportion of affected females would be the square of this, or 0.0001. A classical example is color blindness. Here are data based on about 18,000 school children in Oslo. (Data from Waaler, *Zeit. ind. Abst. Vererb.* 45: 279. 1927).

Table 20-3

	Males		Females		
	Number	Proportion	Number	Proportion	Expected
Color-blind	725	.0801 = p	40	.0044	$p^2 = .0064$
Normal	8324	.9199	9032	.9956	
Total	9049		9072		

As expected, the number of affected females is much less than the number of affected males. Yet the quantitative agreement is not good. There were only 40 affected females when .0064 x 9072, or 58.2 would be expected. The difference is statistically significant.

The explanation is that there are two loci on the X chromosome, one causing red blindness and one causing green blindness. The two loci are complementary, so a female with one gene of each type has normal vision. Hence the frequency of color-blind females is less than expected on the assumption that there is only a single locus. When a more careful analysis is done, taking this into account, the agreement of observation and expectation is very good.

As expected on this hypothesis, occasionally a woman with normal vision has two kinds of color-blind sons, red blind and green blind. That there are two distinct loci has been demonstrated by finding recombinants, although the two genes are quite close together.

PROBLEMS

1. In purebred Holstein cattle about 1 calf in 100 is spotted with red rather than black, red being due to a recessive gene. What is the proportion of red genes in the population? What proportion of the black population is homozygous and what heterozygous?
2. In humans about 1 man in 12 is color-blind, color blindness being due to a sex-linked recessive factor. What proportion of women would be expected to be color-blind?
3. Among 6000 shorthorns chosen at random from the herd-book, 516 were white, 2856 were red, and 2628 roan. Are these roughly consistent with the assumption that roan is due to the heterozygous condition and that mating has been at random for this pair of genes?
4. In man the two genotypes D/D and D/d are Rh+ and d/d is Rh–. The frequency of Rh– people in the population is about 16/100. What proportion of the children in a series of families in which one parent is Rh+ and the other Rh– would be expected to be Rh+? What proportion of the children from a series where both parents are Rh+ would be expected to be Rh–?
5. Given that the frequencies of the blood group genes O, A, and B are .67, .26, and .07 in the United States, reconstruct the frequency data in the first table in Chapter 19.
6. From the data in Chapter 19, what is the frequency of the *cde* chromosome? From the frequency of *CDe*/*CDe* homozygotes, compute the frequency of the *CDe* chromosome. From these two results predict the frequency of the genotype *CDe*/*cde*, and compare with the data.
7. Island A in the South Pacific has 1% albinos in the population. Island B has 4%. Recall that albinos result from a homozygous recessive gene, and assume that both populations are of equal size and have married at random with respect to the albino gene. Due to a geological upheaval these two islands become connected by an isthmus, and the two populations become a single randomly mating unit. What will be the frequency of the albino gene in the combined population? What proportion of the combined population will be albino next generation? How does this compare with the average frequency of albinos in the two populations before the merger? (If you like algebra, prove that it will always be true that combining two populations will give a lower incidence of homozygous recessives than the average of the two original populations.)

as been demon-
genes are quite

) is spotted with
ssive gene. What
pulation? What
ygous and what

color blindness
at proportion of

from the herd-
2628 roan. Are
that roan is due
ng has been at

Rh+ and *d*/*d* is
ulation is about
eries of families
Rh− would be
children from a
expected to be

genes *O*, *A*, and
reconstruct the

ency of the *cde*
homozygotes,
ne. From these
otype *CDe*/*cde*,

the population.
a homozygous
ons are of equal
t to the albino
islands become
ations become a
frequency of the
at proportion of
generation? How
of albinos in the
e algebra, prove
populations will
ssives than the

---

8. Using the data in problems 4 and 5, and knowing that the frequency of the M factor is .56, compute the frequency of persons in the population that are O, MN, and Rh+.

9. Of 359 American Navajo Indians, 87% had blood group O. Compute the 95% confidence limits for this estimate.

10. In humans, baldness is due to a dominant gene that is expressed only in males; females are nonbald, irrespective of genotype. Assume that the frequency of the gene for baldness is p, and that marriage is random with respect to this gene.
    a. What proportion of the sons would be bald from a large series of families where the father was nonbald?
    b. What proportion of the sons would be nonbald from a large series of families where the father was bald?

21

INBREEDING

INBREEDING INCREASES HOMOZYGOSITY

Inbreeding means mating between individuals that are more closely related than the average pair chosen at random from the population. In order to determine the consequences of inbreeding we shall first consider an extreme case, the closest possible inbreeding, in which every individual is self-fertilized.

In a population made up of *A/A, A/a,* and *a/a* genotypes, all self-fertilized, the progeny of the *A/A* and *a/a* types will all be like their parents. But the progeny of the heterozygous type will be on the average 1/4 *A/A*, 1/2 *A/a*, and 1/4 *a/a*. Hence in the next generation the proportion of heterozygous individuals will be only 1/2 as large as it was before and the two homozygous classes will be increased accordingly. If self-fertilization is continued the same thing will happen the next generation; the proportion of heterozygosity is again reduced by 1/2. This is illustrated in the following table which starts with a randomly mating population with gene frequencies p and q each equal to 1/2, and shows the proportions of the three genotypes in successive generations of self-fertilization.

Generation	*A/A*	*A/a*	*a/a*
1	1/4	1/2	1/4
2	3/8	1/4	3/8
3	7/16	1/8	7/16
4	15/32	1/16	15/32

Notice that the gene frequencies are not changing; p and q are still equal to 1/2. The only change is the relative number of homozygotes and heterozygotes. As a general rule we can state that, regardless of the gene frequency, with self-fertilization the heterozygous proportion is reduced by 1/2 each generation.

As a second example consider the consequences of successive brother-sister mating. The computations are more complex and the reduction in heterozygosity each generation is not a simple fraction, but the same general result occurs as with self-fertilization; there is a similar, but less rapid, reduction in heterozygosity. This is illustrated in the following table, again starting with a randomly mating population with p and q each equal to 1/2.

Generation	A/A	A/a	a/a
1	1/4	1/2	1/4
2	5/16	3/8	5/16
3	11/32	5/16	11/32
4	24/64	8/32	24/64

Other forms of less close inbreeding lead to similar results, but at a slower rate. In general the effect of any system of inbreeding is increase of homozygosity and decrease of heterozygosity. Whatever good or harmful effects inbreeding has are due to this. If harmful recessives are in the stock, inbreeding will make them homozygous and hence have a harmful result; if the stock contains desirable hidden recessives inbreeding is useful. Since almost all populations of plants and animals contain some hidden detrimental recessive genes, inbreeding almost always results in some loss of vigor. However, it is sometimes possible to conteract this with selection of the strongest individuals each generation, and the uniformity that goes with homozygosity may be very useful to the animal or plant breeder.

THE COEFFICIENT OF INBREEDING, F

The most widely used measure of inbreeding is Wright's **Coefficient of Inbreeding**, designated by the letter F. It is a measure of the proportion by which heterozygosity has been reduced by inbreeding.

If an individual is inbred, his parents must have been related; if they are related, then they must have one or more ancestors in common. An inbred individual is on the average more homozygous because he is likely to receive an identical gene from each parent, a gene which each parent in turn received from the common ancestor. We shall define F, the inbreeding coefficient, as the probability that two allelic genes in an individual (or in the two gametes that gave rise to the individual) are identical by descent, i.e., both are descended from the same gene in some common ancestor.

Consider now a single locus with two alleles, A_1 and A_2, in relative proportions p_1 and p_2. The probability that two uniting gametes will contain identical genes derived from the same gene in an ancestor is, by definition, F. Thus an individual has a probability, F of being homozygous for such a gene, which may be either A_1 *or* A_2. Since p_1 and p_2 represent the relative frequencies of the two alleles, he has a probability p_1F of being A_1/A_1 and p_2F of being A_2/A_2. The individual also has a chance 1-F that the two genes at this locus did *not* come from the same gene in a common ancestor. In this case he has a chance of p_1^2 of being A_1/A_1 and p_2^2 of being A_2/A_2, as in any randomly mating population. Putting all this together we have the following frequencies of the three genotypes in a population with a coefficient of inbreeding equal to F.

Genotype	Frequency
A_1/A_1	$p_1^2(1-F) + p_1F$
A_1/A_2	$2p_1p_2(1-F)$
A_2/A_2	$p_2^2(1-F) + p_2F$

More generally, for multiple alleles,

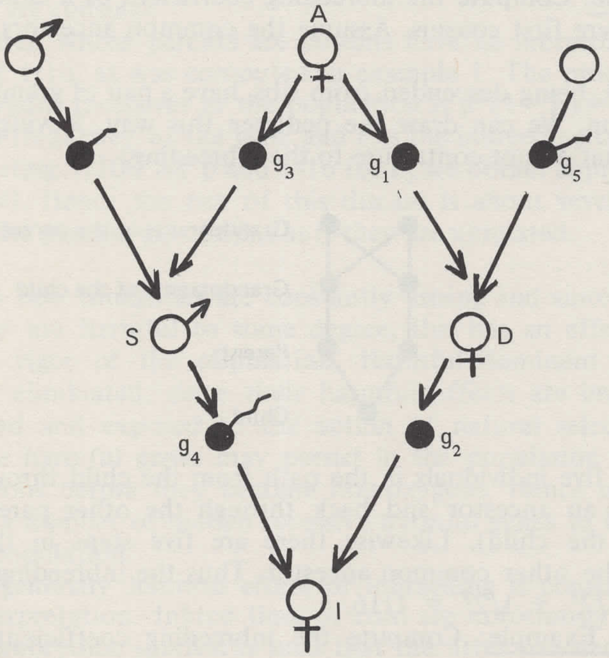
Homozygote	A_iA_i	$p_i^2(1-f) + p_iF$
Heterozygote	A_iA_j	$2p_ip_j(1-F)$

Notice that when $F = 0$ (i.e., random mating) the values reduce to the usual Hardy-Weinberg formula. When $F = 1$, the maximum inbreeding, the population is completely homozygous.

By inbreeding, the proportion of heterozygotes has changed from $2p_1p_2$ to $2p_1p_2(1\text{-}F)$, or has been reduced by a fraction F. This is generally true, regardless of the number of alleles or number of loci: in a population of inbreeding coefficient F, the proportion of heterozygous loci is reduced by a factor F, compared with what it would be if mating had been random.

COMPUTATION OF F FROM PEDIGREES

In the diagram below, suppose that gamete g_1 carries a certain gene. The probability that gamete g_2 carries a descendant of this same gene is 1/2, for the gene in g_2 has an equal chance of having come from the egg g_1 or the sperm g_5. In a similar way the two gametes g_4 and g_3 have a probability of 1/2 of both containing the same gene. Now, call the two alleles in ancestor A, A_1 and A_2. The two gametes g_3 and g_1 may contain A_1 and A_1., A_2 and A_2, A_1 and A_2, or A_2 and A_1. In the first two cases the genes are identical. In the second two cases the genes are different, unless ancestor A is inbred, in which case the two genes A_1 and A_2 may themselves be descended from the same gene in some more remote ancestor. The probability of this is, by definition, the inbreeding coefficient F_A of the individual A. Thus the probability that g_3 and g_1 contain identical alleles is $1/2 + F_A/2$, or $(1 + F_A)/2$.



Putting all these together, the probability that g_2 and g_4 contain identical alleles is the product of the three probabilities. Hence the inbreeding coefficient of individual I is

$$F_I = 1/2 \times 1/2 \times (1+F_A)/2 = (1/2)^3 (1+F_A)$$

By extension of this principle, the contribution of any common ancestor, A, to the inbreeding coefficient of an individual, I, is $(1/2)^n (1 + F_A)$ where n is the number of individuals in the path leading through one parent to the common ancestor and back through the other parent. Finally, there may be more than one common ancestor, or more than one path to the same ancestor. The total probability of individual I having a pair of identical alleles is the sum of all the ways in which this could happen, so we sum the contribution for all paths through all common ancestors. Symbolically, we write

$$F_I = \Sigma \; [(1/2)^n (1 + F_A)] \qquad (21\text{-}1)$$

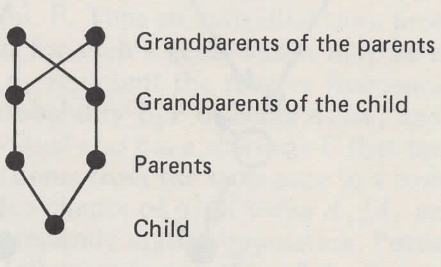
where n is the number of individuals in a particular path, not counting individual I; F_A is the inbreeding coefficient of the common ancestor in this path; and Σ means to add up all the terms, one for each path.

If the genes being considered are sex-linked, only the females in the path are counted, and the path is ignored entirely if it has two consecutive males.

We can also define the **Coefficient of Kinship** (or **Consanguinity**) of two individuals, J and K, as F_{JK}. This is the probability that a randomly chosen pair of alleles, one from each individual, are identical by descent. Obviously, if I is the progeny of J and K, then $F_I = F_{JK}$.

Example: Compute the inbreeding coefficient of a child whose parents were first cousins. Assume the common ancestors are not inbred.

Cousins, being descended from sibs, have a pair of grandparents in common. We can draw the pedigree this way, leaving out all persons that do not contribute to the inbreeding:



There are five individuals in the path from the child through one parent to an ancestor and back through the other parent (not counting the child). Likewise there are five steps in the path through the other common ancestor. Thus the inbreeding coefficient is $1/2^5 + 1/2^5 = 1/16$.

Second Example: Compute the inbreeding coefficient of in-

that g_2 and g_4
hree probabilities.
s
)

n of any common
n individual, I, is
iduals in the path
ncestor and back
e more than one
ne same ancestor.
pair of identical
ld happen, so we
mmon ancestors.

(21-1)

ticular path, not
cient of the com-
up all the terms,

ly the females in
rely if it has two

hip (or **Consan-**
K· This is the
, one from each
I is the progeny

of a child whose
ncestors are not

of grandparents
leaving out all

ne parents

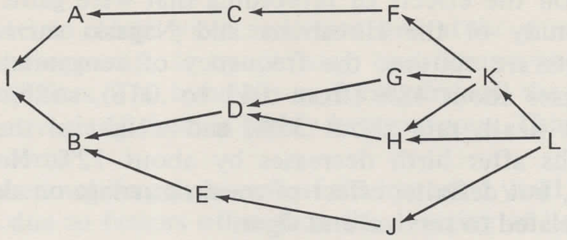
ne child

d through one
er parent (not
s in the path
reeding coeffi-

fficient of in-

---

dividual I in the pedigree below, given that the inbreeding coefficient of K is 1/4, and L is zero.



The contribution of the path	ACF$\underline{K}$GDB	is	$1/2^7(1+1/4) = 5/512$
	ACF$\underline{K}$HDB		$1/2^7(1+1/4) = 5/512$
	ACFK$\underline{L}$JEB		$1/2^8 = 1/256$
	ACFK$\underline{L}$HDB		$1/2^8 = 1/256$
	Total		$=14/512=.027$

In the analysis above, the common ancestor has been underlined. Notice that in any particular path it is irrelevant whether any individual is inbred except the common ancestor.

There are two other things to keep in mind: (1) an individual never occurs more than once in a single path; (2) a path never zigzags; on the way to a common ancestor it always goes from offspring to parent, on the way back, always from parent to offspring. For example, the path ACFKGDHLJEB is not permitted.

Third Example: From the previous chapter the frequency of the gene for phenylketonuria is about 1/100. The risk of having a child with the disease when the parents are unrelated is the square of this, or about 1/10,000. How much is this risk enhanced if the parents are cousins?

Children whose parents are cousins have an inbreeding coefficient of 1/16, as was computed in example 1. The proportion of homozygous recessives in the population is $p^2(1 - F) + pF$, where p is the frequency of the gene and F is the inbreeding coefficient. Substituting 1/100 for p and 1/16 for F, we obtain approximately 7/10,000. Hence the risk of this disease is about seven times as high if the parents are cousins as if they are unrelated.

EMPIRICAL DATA ON INBREEDING AND ESTIMATES OF THE AMOUNT OF HIDDEN GENETIC DAMAGE DUE TO RECESSIVE GENES

Since new mutations are constantly arising and since the great majority are harmful to some degree, this has an effect on the average vigor of the population. Harmful dominant genes are quickly eliminated, since their harmful effects are immediately expressed and exposed to the action of natural selection. But recessive harmful genes may persist in the population for many generations before they become homozygous. Hence one would expect a number of hidden recessive harmful genes to be present in any population.

The generally harmful effect of inbreeding is consistent with this interpretation. Inbred lines of corn are notoriously weak. In swine, embryonic survival is such that the litter size decreases by about one pig for each increase of 0.1 in the inbreeding coefficient. Considerable information is available for man, largely due to data on the effects of inbreeding that were gathered in Japan in the study of the Hiroshima and Nagaski survivors. When the parents are cousins, the frequency of congenital malformations increases about 48% (from .011 to .016), stillbirths about 25%, infant death rate about 35%, and weight of the child about 9 months after birth decreases by about 12%. Hence, there is a slight, but definite, effect of cousin marriage on almost all characters related to survival and vigor.

We can use data from consanguineous marriages to estimate how much genetic weakness is carried in the form of hidden recessive genes. Data gathered from church records in rural France before World War II yielded the following results:

Deaths before adulthood (including stillbirths)

Parents unrelated	.12
Parents first cousins	.25
Excess from cousin marriage	.13

Recall that the inbreeding coefficient for a child of a cousin marriage is 1/16. This means that 1/16 of the genes that would otherwise be heterozygous are made homozygous by cousin marriage. Multiplying 0.13 by 16 gives approximately 2. Hence if a person could be made completely homozygous (e.g., if he were a Paramecium and could undergo autogamy) he would have twice as many homozygous harmful recessive genes as would be required to cause infant or childhood death. Since a homozygous individual is equivalent to a doubled gamete, we can say that there is the equivalent of two lethal genes per gamete. Thus the typical diploid individual carries four **lethal equivalents** in the form of hidden recessive genes. We say lethal equivalents instead of lethal genes, because there may be 8 genes each causing 50% probability of death, for example. There is no way to distinguish among these possibilities in man, although in Drosophila we know that about 2/3 of the reduced viability from inbreeding is caused by lethal genes and the rest by cumulative effects of a large number of mildly deleterious genes.

The data on which these calculations are made were gotten some time ago when living conditions were harder than they are now. Current data from Japan show values less than half as large. So, apparently a gene that caused death in France at that time causes less than 50% chance of death now in Japan, because of higher living standards, better medical care, and improved public health conditions.

CROSSING INBRED LINES FOR UNIFORMITY AND HIGH YIELD

Inbreeding, by increasing the homozygosity, brings out the detrimental genes that were previously concealed by heterozygosis with more favorable dominant alleles. It is usually impossible, even by the most rigid selection, to rid an inbred line of all detrimental recessives, so a decline in vigor, yield, and survival are almost inevitable consequences of inbreeding. At the same time inbreeding increases the uniformity by decreasing the number of segregating loci.

One way to attain a high yield and vigor, and at the same time have genetic homogeneity, is by crossing two homozygous (or nearly so) inbred lines. The hybrid is uniform, since all the plants are identical, yet each of them is highly heterozygous. This is the basis of hybrid corn.

Evidence is accumulating that some of the increased yield in the best hybrids is due to factors other than the covering of harmful recessives. There are some *overdominant* loci at which the heterozygote, A/a, is superior to either homozygote, A/A or a/a. How important such loci are for economic characters like yield in corn is unknown.

By inbreeding it is possible to get corn strains with desirable attributes, such as proper ear shape, sugar content, color, and resistance to disease. Then two inbred strains are crossed to get the increased yield that results from heterozygosity. The only difficulty with this scheme is that the corn seed that is sold for planting is grown on inbred plants; this cuts down the yield and makes the seed expensive. This difficulty is circumvented by using four-way crosses. Inbred lines A and B are crossed to produce hybrid AB; C and D are crossed to produce CD; then the two hybrids are crossed. The resulting four-way hybrid is just about as high yielding as the single cross, and though somewhat more variable, is satisfactory for commercial purposes.

Hybrid corn is produced from specially bred hybrid seed, and the seed has to be purchased anew each season. The following data show the decline in yield when a field of hybrid corn is allowed to pollinate at random and then these seed are planted. A three-way cross is a cross between a hybrid and an inbred, AB x C.

Table 21-1
Yield in bushels/acre (Data from N. P. Neal)

	Inbred lines from which hybrid derived (F = 1)	Hybrid (F = 0)	Randomly pollinated from hybrids	Percent decline	F after random pollination
Single cross	23.7	62.8	44.2	47.6	.50
Three way	23.8	64.2	49.3	36.8	.375
Four way	25.0	64.1	54.0	25.8	.25

Notice that the decline in yield is almost exactly what would be expected from the inbreeding coefficient.

RANDOM GENE FREQUENCY DRIFT IN SMALL POPULATIONS

In a very large population, if mating is at random, it is very unlikely that two mates are close relatives. On the other hand, if a population is small, and especially if it has been small for a long time, many individuals will be related to one another. Thus random matings are likely to be consanguineous. We would expect, then, that a small population, even when mating is random, would have effects very much like inbreeding in a large population.

To illustrate this, I again use the gene pool device. Suppose there are N diploid individuals in the parent generation, and that each of these contributes its two genes to the gene pool. To get the offspring generation we choose genes at random with replacement after each draw. The probability that we draw the same gene twice is 1/2N, since there are 2N genes in the pool. The chance that we draw two different genes is, of course, 1-1/2N. However, the probability that the two different genes are in fact the same gene tracing back to some more remote common ancestry is the inbreeding or kinship coefficient of the parent generation. Since mating is random, the probability of identity of two genes in the same individual is the same as that of two genes in two different individuals.

If we let F_t represent the inbreeding coefficient—that is, the probability of identity—in the t-th generation, then

$$F_t = \frac{1}{2N} + (1 - \frac{1}{2N})F_{t-1}$$

The meaning will be more transparent if we rewrite this as

$$1 - F_t = (1 - \frac{1}{2N})(1 - F_{t-1})$$

Now, recall that 1-F is proportional to the heterozygosity, so if we let H stand for heterozygosity,

$$H_t = (1 - \frac{1}{2N})H_{t-1} \qquad (21\text{-}2)$$

or, in words, the heterozygosity decreases each generation by a fraction equal to 1/2N.

Our gene pool assumes complete random combination of gametes. We are usually most interested in organisms like ourselves that do not have self-fertilization. I won't explain it here, but if N is at all large, this makes very little difference. The effect of prohibiting self-fertilization is to change 1/2N to approximately 1/(2N + 1).

This result, first discovered by Sewall Wright, measures random mating in a finite population on the same scale as nonrandom mating in an infinite population. For example, you can compute that the kinship coefficient of second cousins is 1/64, and therefore the reduction in heterozygosity from second cousin marriage is 1/64. Thus a population of size 32 mating at random has the same heterozygosity consequences as a second cousin mating.

There is one tricky aspect of this change in heterozygosity in a finite population. The genotypes within the population are in

Hardy-Weinberg proportions, since mating is at random. What is happening is that there are random changes in the gene frequencies. The smaller the population, the greater these random changes will be. You might think that these are equally likely to be up or down, and this is true. Yet, on the average the gene frequency changes so that the quantity $2p_A p_a$ is decreased next generation to 1-1/2N of its present value.

The size of the breeding population in nature may be smaller than it first seems. Only mating adults count, of course. The population may be broken up into smaller units within which random mating occurs. There may be a discrepancy in the numbers of males and females. Some individuals may have many more progeny than others. The population may fluctuate in size, especially between winter and summer, and chance may play a large part in determining which of the large number of individuals in a summer population survive the winter to become the parents of next summer's population. For all these reasons, the *effective* population number may be considerably smaller than the census number.

As a numerical example, suppose an island population starts with 20 shipwrecked people (after the manner of the mutineers from the "Bounty" who, along with some Tahitian men and women, founded the population of Pitcairn Island). To simplify matters, suppose that these are equally divided between men and women and that the population size does not change, remaining steadily at 20 adults. After five generations the average heterozygosity would have decreased about 12%. This is about the equivalent of an uncle-niece marriage. Any recessive gene has about a 12% chance of being expressed through homozygosis. We would expect a much higher frequency of recessive disorders than in a large population.

There is also a qualitative feature of an isolated population. In this population no gene frequency can have been less than 1/40, since the founders were 20 diploid individuals. Mutants that are very rare in a large population, in this population would either be nonexistent (most of the time) or fairly common with a fair probability of being homozygous. It is no surprise that medical geneticists studying rare conditions like to study populations that for geographical, cultural, or religious reasons have remained small and isolated. Diseases that are ordinarily very rare are found in sufficiently high frequency for effective comparative study. Each population has its own special kinds of mutant genes, different from those of other isolated populations, and being there by the accident of having been present in one of the founders. There are numerous anecdotes of genes in isolated communities that can be traced back to a single ancestor. Evolutionists refer to this phenomenon as the **founder effect**.

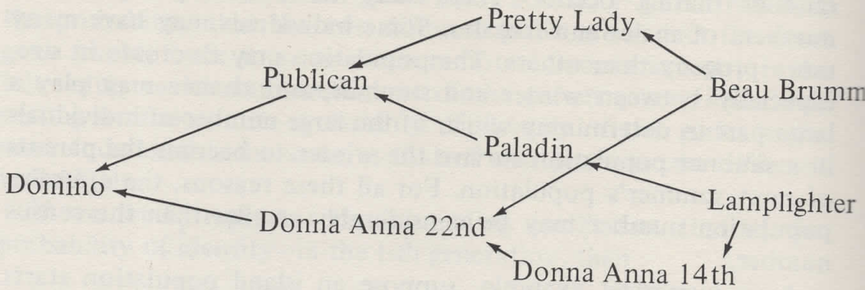
PROBLEMS

1. If the average proportion of heterozygosity in a population is 1/3, what would it be after two generations of self-fertilization?

2. What would be the inbreeding coefficient of a child whose parents were second cousins (i.e., the children of first cousins)?

3. What is the inbreeding coefficient of a child whose parents were half sibs? Uncle and niece? Double first cousins? Society has quite different attitudes toward marriage of people of these relationships. Is there any genetic basis?

4. Determine the coefficient of inbreeding of the Hereford bull, Domino, from this pedigree.



5. What would be the inbreeding coefficient of Domino if Lamplighter has an inbreeding coefficient of 1/4?

6. By computing F, verify the result shown at the beginning of this chapter that the proportion of heterozygosity in the offspring of a brother-sister mating is 3/8 when the gene frequency is 1/2.

7. A major problem in growing hybrid corn seed is preventing self-fertilization in the single cross plants that are to be used as female parents of the double cross seed. This is accomplished by the expensive operation of removing the pollen-bearing tassels from these plants. However, in corn there is a cytoplasmic factor causing sterility of the male part of the plant (as was mentioned in Chapter 17). Plants with this factor are male-sterile if their genotype is *A/- b/b*, but with all other genotypes the plant is male-fertile. Using these genes (and their alleles) and the cytoplasmic factor, work out a system whereby one of the single cross strains will be male-sterile, but the double cross is male-fertile.

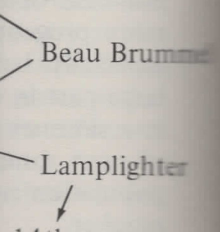
8. A group of four inbred lines average 40 bushels per acre. The four-way hybrid averages 100 bushels. What would you predict for the yield of randomly mated progeny of the four-way hybrids?

9. Regard *MS, Ms, NS,* and *Ns* as four alleles with frequencies p_1, p_2, p_3, and p_4 respectively.
   a. What fraction of a randomly mating population would react with all four antisera (anti-M, anti-N, anti-S, and anti-s)?
   b. What fraction of the population would be MS/Ms?

child whose
dren of first

vhose parents
usins? Society
of people of

Hereford bull,



14th
f Domino if

beginning of
gosity in the
en the gene

is preventing
to be used as
accomplished
ollen-bearing
re is a cyto-
the plant (as
is factor are
ith all other
es (and their
ystem where-
erile, but the

per acre. The
you predict
he four-way

quencies p_1-

ation would
anti-S, and

Ms?

---

c. What fraction of people would be MS/MS in a hypothetical population in which everybody married his cousin?

10. Regard each of the Rh triplets as an allele with frequencies D_{CDE}, p_{CDe}, etc.
    a. What fraction of a randomly mating population would be heterozygous for all three letters?
    b. What fraction would be *cde*/*cde* in a hypothetical population where everybody married his cousin?

11. Show from the formula $H_t = (1 - 1/2N)H_{t-1}$, that $H_t=(1-1/2N)^t H_O$. Use this to calculate (if you have access to a pocket computer) the amount of heterozygosity, relative to the original amount, after 100 generations in a population of 50. (If you like math, you may want to show that the answer is approximately 1/e.)

12. Verify the F values in Table 21-1 on page 175.

SELECTION

22

INTRODUCTION

If one genotype survives more often or produces more offspring than another, the gene frequency will change. Change of gene frequency by such differential survival and reproduction is called selection. Selection may be (1) artificial, as practiced by the livestock or plant breeder who saves the best individuals to use as breeding stock, or (2) natural, as occurs in nature when the weaker forms die or fail to reproduce.

SELECTION AGAINST A DOMINANT GENE

Getting rid of an unwanted or disadvantageous dominant gene in an animal or plant population is quite simple. If no individuals possessing the gene reproduce, the next generation will be made up entirely of homozygous recessives (except for the possibility of mutations). Thus, the livestock or plant breeder has no serious problem in ridding his stocks of occasional undesired dominant factors. Of course, the breeder may not be able to afford to cull all the animals with the dominant phenotype, or if the gene has incomplete penetrance some animals carrying the gene may not be detected. In such cases the rate of progress in ridding the population of the undesired factor is proportionately slower.

SELECTION AGAINST A RECESSIVE GENE

Selection against a recessive factor is quite different. Removal of homozygous recessives from the population still leaves many recessive genes hidden in heterozygotes. As the recessive gene becomes rare the rate of progress in reducing its frequency becomes slower. This can be shown quantitatively as follows:

Starting with a population in the proportions p^2 A/A, $2pq$ A/a, and q^2 a/a, where $p + q = 1$, we remove the homozygous recessives. The only way in which recessive homozygotes can occur next generation is from the mating of two heterozygotes, which should produce on the average 1/4 a/a offspring. The proportion of heterozygotes among those remaining after removal of the a/a genotypes is

$$\frac{2pq}{p^2 + 2pq} = \frac{2q}{1 + q}$$

The probability of a homozygous recessive progeny next generation is the probability that the mother is heterozygous times the probability that the father is heterozygous times 1/4, or

$$\frac{1}{4}\left(\frac{2q}{1+q}\right)\left(\frac{2q}{1+q}\right) = \left(\frac{q}{1+q}\right)^2$$

So, if q_0^2 is the initial frequency (generation zero), the frequency after one generation of selection (generation 1) is

$$q_1^2 = \left(\frac{q_o}{1+q_o}\right)^2$$

After two generations of selection (generation 2) the frequency is

$$q_2^2 = \left(\frac{q_1}{1+q_1}\right)^2 = \left(\frac{q_o}{1+2q_o}\right)^2$$

and after n generations the frequency is

$$q_n^2 = \left(\frac{q_o}{1+nq_o}\right)^2 \qquad (22\text{-}1)$$

In Holstein and Aberdeen-Angus cattle there are still occasional calves born with red coloring instead of the desired black, despite many years of rigorous selection. At present something like 1 calf in 100 is red. This means that with the present practice of culling all reds the frequency next generation will change to 1 in 121. In 10 generations of selection the proportion will be reduced to about 1 in 400. Unless there is some way of detecting and eliminating the heterozygous carriers, selection against rare recessive genes is a very slow process.

SELECTION AGAINST A SEX-LINKED RECESSIVE GENE

If there is complete elimination of homozygous recessive females and recessive males, then after the first generation there will be no more homozygous recessive females. From that time on, all the recessive phenotypes will be males, who in turn came from heterozygous mothers. A heterozygous female may transmit the recessive gene to a son or to a daughter; in the former case it is eliminated, in the latter it is retained by the population. Hence the proportion of heterozygous females is reduced by half each generation. Likewise the proportion of recessive males will be reduced by half each generation. The proportion of affected males born each generation will be half the proportion of heterozygous females the previous generation.

PARTIAL SELECTION

Much of the time natural (or artificial) selection is not as intense as in the examples we have been considering. Instead of one genotype or phenotype being prevented from leaving *any* progeny in the next generation, one phenotype may be only a little less effective in survival or reproduction than others. And the situation may be complicated by the fact that dominance may not be quite complete.

As the intensity of selection varies, the effectiveness of selection in changing the population varies too. Intense selection can result in quite rapid changes—except, as we have just learned, when the gene being selected for or against is recessive and rare. Even mild selection can spread a desirable gene through the population in time, but the time may be exceedingly long. As an allele approaches neutrality in terms of its selective value to the organism, selection becomes less and less important compared with other processes in determining the gene frequency.

We shall use **fitness** for the combined ability to survive and reproduce. If we start out with 100 zygotes of each of two types, and the first group leaves 100 descendants in the next generation while the second leaves 95, we say that the **relative fitness** of the second group is 0.95 or that the second group is 5% less fit. Again we assume that mating is random and use the gene pool model. Assume that each parental genotype contributes to the gene pool in proportion to its fitness, and then pairs of genes are chosen at random from this pool.

Of course this is not the way a real population behaves. The fertility of an individual may not be a constant, but be dependent on the mate; for example, in a strictly monogamous species, if one of the two mates is sterile, both are. Furthermore, the relative fitness of a genotype is likely not to be a constant, but rather to depend on what other genotypes are present, the degree of crowding, and many other factors. However, our simple model has the merit of being manageable by simple algebra. It gives answers that are a very good approximation to reality when selection is weak, and it also gives us qualitative insights into the way selection works.

Let A stand for the gene favored by selection and A' for the other allele. To avoid too many subscripts I will use p for p_A and q for $p_{A'}$. Remember that $q = 1 - p$. The relative fitnesses are taken to be 1, $1 - hs$, and $1 - s$. If $s = .05$, then the $A'A'$ type is 95% as fit as AA. The letter h is a measure of dominance. When $h = 0$, AA and AA' have the same fitness, so A is dominant; when $h = 1$, A is recessive; and when $h = 1/2$ there is no dominance either way. Thus, h can take any value between 0 and 1; in fact it may have values outside this range if there is overdominance. I'll consider such examples later. We let s stand for selection for mnemonic convenience, and h for heterozygous effect. Then the model is:

Table 22-1

Genotype	*AA*	*AA'*	*A'A'*	Total or average
Relative fitness	1	$1 - hs$	$1 - s$	
Frequency	p^2	$2pq$	q^2	1
Relative contribution to reproductive gene pool	p^2	$2pq(1 - hs)$	$q^2(1 - s)$	$\overline{w}$

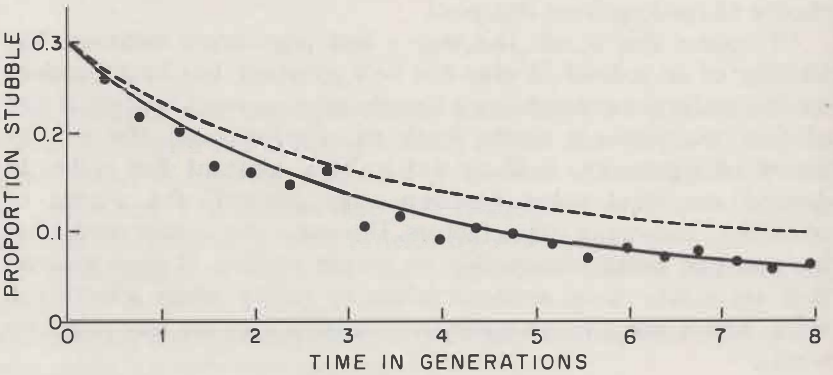
$$\overline{w} = p^2 + 2pq(1 - hs) + q^2(1 - s) = 1 - sq(q + 2ph)$$

The proportion of *A* genes next generation is

$$p' = \frac{p^2 + pq(1 - hs)}{\bar{w}} = \frac{p(1 - qhs)}{\bar{w}} \qquad (22\text{-}2)$$

By using this formula we can compute the expected gene frequency generation after generation. Knowing the gene frequency, we can compute the genotype frequency from the Hardy-Weinberg relations.

An example showing that these equations work very well in real populations is shown below. The data come from a Drosophila population cage studied by Y. J. Chung. The graph shows the proportion of mutant flies in successive generations.



There were three genotypes, +/+ (normal), +/*Sb* (stubble bristles), and *Sb*/*Sb* which is lethal in the larval stages and hence not seen. The average age of reproduction is about 2.5 weeks. A sample of 200 adults was removed every week and the numbers of normal and stubble-bristled flies were counted. The graph shows the proportion of stubble. The data in the figure are the average of four cages started simultaneously. The dotted line is the expectation if the mutant behaves as a recessive lethal. Clearly, this does not fit. The solid line is for $s = 1$ and $h = .12$.

The real population has eggs, larvae, pupae, and adults all coexisting in a cage. The generations are overlapping. The environment is relatively constant, but certainly not absolutely so; for example, the population size fluctuated considerably. Selection is very strong, being complete for one genotype and 12% in the heterozygotes. Yet, a simple model based on random selection from a gene pool at discrete time intervals provides an excellent fit to the data.

Using equation 22-2, I have computed the gene frequency generation after generation. The results for selection in favor of a dominance ($h = 1/2$) are shown on opposite page. The table gives the number of generations required to change the gene frequency a specified amount when the selection intensity, s, is 0.01.

Table 22-2

Gene frequency From	To	Dominant favored (h = 0)	Recessive favored (h = 1)	No dominance (h = 1/2)
.001	.01	232	90,231	462
.01	.10	250	9,240	480
.10	.25	132	710	220
.25	.50	177	310	220
.50	.75	310	177	220
.75	.90	710	132	220
.90	.99	9,240	250	480
.99	.999	90,231	232	462

There is a reciprocal relation between the intensity of selection, s, and the time required for a specified gene frequency change (provided s is small). Thus to find the time required when $s = 0.1$ instead of 0.01, divide each of the numbers by 10. For example, if you would like to know how long it takes to change the frequency of a favored dominant gene from .01 to .5 when $s = .01$, you note first that the time required when $s = .01$ is 132 + 177, or 309 generations. The time when s is 10 times as large is 1/10 as many generations or 31.

SELECTION BALANCED BY MUTATION

The results of the previous section have shown that consistent selection changes the gene frequencies, sometimes rapidly, more often slowly. Eventually, although it may take a very long time, the favorable genes replace the less favorable. Yet, as we see whenever we study a natural population of any species, there are always a number of deleterious genes. One explanation of these is that they arise continuously by mutation.

We would expect in a natural population, including the human population, that new deleterious genes occur through mutation but they are eliminated by natural selection. Eventually these two processes should come to a balance when the rate of occurrence of new mutants equals the rate of extinction through natural selection. Let's consider this balance analytically.

As a first case consider a recessive mutant, with the same model as before. In this case $h = 0$.

Genotype	*AA*	*Aa*	*aa*	
Frequency	p^2	$2pq$	q^2	$p + q = 1$
Relative fitness	1	1	$1 - s$	

The *a* gene, being deleterious from the standpoint of survival and fertility, will tend to decrease. We now introduce mutation into the picture by assuming that the *A* gene mutates to *a* at a rate u per gene per generation.

The frequency of the A allele at the beginning of the next generation (see equation 22-2 for $h = 0$) is

$$p' = \frac{p^2 + pq}{\bar{\bar{w}}}\,(1 - u) \;=\; \frac{p(1-u)}{\bar{\bar{w}}}$$

The term $1 - u$ is introduced because we wish to count only those *A* genes that have not mutated. The rest of the formula is the same as we used earlier, taking h as 0. In this case $w = p^2 + 2pq + q^2(1\text{-}s) = 1 - sq^2$.

Now, when the two processes of origin of new genes by mutation and their extinction by selective elimination come into balance, the gene frequency will not change. We specify this by setting $p' = p$. Making use of this relationship (don't forget that $p + q = 1$), the equation reduces to

$$q^2 = u/s, \text{ or } q = \sqrt{u/s} \tag{22-3}$$

The rule is easy to remember because it is so reasonable: the proportion of recessive homozygotes at equilibrium is the ratio of the mutation rate to the selective disadvantage of the mutant homozygote.

What if the mutant is dominant or partially dominant? The equations are more complex, but the principle is the same. We will again use the earlier model where h is a measure of dominance.

Genotype	*AA*	*AA'*	*A'A'*
Frequency	p^2	2pq	q^2
Fitness	1	1 - hs	1 - s

Equation 22-2 is modified to include the term $1 - u$. Thus

$$p' = \frac{p(1 - qhs)}{\bar{w}}(1 - u), \text{ where } w = 1 - sq(q + 2ph)$$

On setting $p' = p$, we obtain the quadratic equation

$$s(1 - 2h)q^2 + hs(1 + u)q - u = 0,$$

and we can solve this to get the value of q for any fixed values of h, u, and s. Notice that the example of a complete recessive falls out as a special case when $h = 0$.

We are mainly interested in cases where h is not close to zero, that is, where there is appreciable selection against the heterozygote. The solution then is given to quite a good approximation by

$$q \simeq \frac{u}{hs} \tag{22-4}$$

Again, this makes sense. The frequency of the gene is given by the ratio of the mutation rate to the selective disadvantage of the heterozygote. If heterozygotes are appreciably selected against, homozygotes are too rare to matter.

Let's now use these equations to understand mutations in natural populations. The best information is for lethal mutations in Drosophila. The spontaneous mutation rate, as we saw in Chapter 15, is about .002 recessive lethals per X chromosome per generation. There are about 1000 loci on the X chromosome, so the mutation rate per locus is (2/1000)/1000, or

$$u = 2 \times 10^{-6}$$

When an autosome is made homozygous, we find about one lethal per four chromosomes. There are about 2000 gene loci on the chromosome, so the lethal gene frequency is (1/4)/2000, or

$$q = 1.25 \times 10^{-4}$$

Since we are discussing lethal mutations,

$$s = 1$$

From equation (22-3) for completely recessive lethals, we expect that $q = \sqrt{u}$, but clearly the data are not in agreement: q is less than 1/10 as large as it should be. So this must mean that the lethals are not completely recessive, or that the population is not mating at random, or that it has not reached equilibrium. The latter two possibilities are largely ruled out by long-time studies of artificial populations in laboratories, which can be followed until they reach an equilibrium and in which there is no opportunity for local inbreeding.

So, accepting that there is some selection against lethal heterozygotes, let's use equation (22-4) to estimate h. From this equation

$$h = u/qs = .016$$

We conclude that the average "recessive" lethal in a natural population has about 1.6% selective disadvantage in the heterozygous state.

Although the information is much less precise, the same thing seems to be true in the human population. Inbreeding causes decreased viability and incidence of recessive diseases, but not as frequently as would be expected if the genes were completely recessive; so, as in flies, there is probably selection against heterozygotes.

We can also use this as a method for computing the mutation rate. A good example is the short-limbed dwarfism, achondroplasia, caused by a dominant gene. A study in Denmark showed an incidence of 10 among 94,075 births. These were all heterozygotes, so the mutant gene frequency is 1/2 the incidence, or 5/94,075. The total number of achondroplastics known was 108, including those who had died, and they produced 27 children. A total of 457 normal sibs produced 582 children. The relative fitness of the dwarfs then is estimated as (27/108)/(582/457), or 0.19. We therefore estimate that hs = .81. Thus u = .81 X 5/94,075 or about 4/100,000. This agrees well with the mutation rate estimated from the number of dwarfs born from normal parents.

I won't bother to derive it, but you might be interested to know that the proportion of males which are affected with an X-linked recessive gene is 3u/s, where s is the selective disadvantage of the affected males.

EQUILIBRIUM BETWEEN OPPOSING SELECTION FORCES

Under the conditions of selection so far discussed, we would expect a favored mutant allele to approach fixation in a population and then be maintained at a high frequency, but kept short of complete fixation by recurrent mutation to other forms of the gene. We might then ask why populations do in fact retain so much genetic heterogeneity. Why, for example, are the A-B-O blood groups alleles kept in our population at intermediate levels of allele frequency? Why are we not entirely O or A or B? Why is the allele for sickle cell hemoglobin, so patently disadvantageous in the homozygote, maintained at a high frequency in many populations? Many other examples will occur to you.

The answers are complex and for the most part not understood. The best understood example is sickle cell hemoglobin. The abnormal hemoglobin, S, provides a condition unfavorable for the malarial parasites in the red blood cell. In an environment that includes endemic malaria, particularly of the fatal falciparum type, individuals with only A type hemoglobin are at a serious disadvantage. At the same time, individuals homozygous for sickle cell hemoglobin are at a disadvantage for another reason—they have sickle cell anemia. The heterozygotes have neither the anemia nor the susceptibility to malaria.

Sickle cell anemia is common in persons of African descent and also is found in some other parts of the world where malaria is, or has been, a major cause of death. The gene, despite the fact that it causes a severe anemia in the homozygous state, is preserved in the population as a protection from malaria (and perhaps for other reasons not understood).

The most revealing way to designate the fitness values in this case is:

Genotype	*AA*	*AS*	*SS*
Frequency	p_A^2	$2p_A p_S$	p_S^2
Relative fitness	1 - a	1	1 - s

We proceed in a now-familar pattern—expressing the gene frequency next generation as a function of the genotypic frequencies and fitnesses this generation. Doing this, we get

$$p'_A = \frac{p_A^2(1 - a) + p_A p_S}{p_A^2(1 - a) + 2p_A p_S + p_S^2(1 - s)} = \frac{p_A(1 - ap_A)}{1 - ap_A^2 - sp_a^2}$$

Letting $p'_A = p_A$ and doing a bit of algebraic manipulation (need I remind you that $p_A + p_S = 1$?) we get

$$p_A = \frac{s}{a + s} \; ; \; p_S = \frac{a}{a + s}$$

Assuming (realistically) that the genotype *SS* is effectively lethal (s = 1) and that the gene frequency in a part of Africa where malaria is common is $p_S = 0.1$, what do you estimate to be the relative fitness of normal individuals compared to the heterozygote?

Answer: $.1 = \frac{a}{a + 1}$, $a = 1/9$, $1 - a = 8/9$

Malaria, in the past and in many parts of the world even today, is a most significant disease. As such, it has played a large role in human evolution. It has also been involved in the establishment of other genetic systems. The global distribution of the X-linked recessive allele for glucose-6-phosphate dehydrogenase deficiency which causes red-cell hemolysis in response to certain drugs or articles of diet (especially a certain kind of bean), and of thalassemia, another type of anemia also coincides quite closely with a map of malaria distribution. These genes, like that for hemoglobin S, bring about resistance to malaria at the price of producing anemia.

There doesn't seem to be any way for a genotype to be both resistant to malaria and free of anemia except by being heterozygous. Perhaps, if left to its own devices, evolution might some time produce such a homozygous genotype, perhaps by the right kind of gene duplication. But, even if the right duplication should occur, it would require many generations of natural selection for it to become the prevailing type. Fortunately, malaria is on the way out as a major cause of human misery.

Is an adaptively superior heterozygote the principal mechanism by which polymorphisms are maintained? This is certainly a plausible mechanism, and in a few cases demonstrated to be the one at work. But in most instances the answer is not known, so I'll mention one other reasonable possibility that has also been demonstrated to work in some instances. This is a situation where genes are favorable when rare, but unfavorable when common. It is easy to see that this would have the postulated effect. A rare gene would tend to increase, so the population would tend to maintain two or more alleles, since if any gene started to decrease in frequency, selection would bring its frequency back up again. Does this make sense physiologically? Yes, at least in some instances. The more similar two individuals are, the more they compete for the same resources (for example, they would have the same food requirements). Thus a rare type would have the advantage of competing mostly with others with slightly different requirements, and with very few who have the same requirements. Much work is needed here before we will know which of these two mechanisms is most important, or indeed if there are others, not yet understood, that are still more important.

It has also been suggested by M. Kimura that many isozyme polymorphisms are essentially neutral from the standpoint of natural selection, and that the polymorphism is simply the result of recurrent mutation.

SELECTION AGAINST HETEROZYGOTES: MATERNAL-FETAL INCOMPATIBILITY

We have seen that selection favoring heterozygotes leads to a stable equilibrium in which both alleles (or more if several alleles are involved) remain in the population. There is an equilibrium value for the allele frequencies and, if for any reason the population departs from this, it will return. It won't surprise you to learn that selection *against* heterozygotes has the opposite effect.

Suppose a population is made up of 1/4 *AA*, 1/2 *Aa*, and 1/4 *aa*. Now assume that the *Aa* class has lower fitness so that it leaves fewer descendants. However, the reproductive gene pool will still contain an equal number of *A* and *a* genes, and with random mating the zygotic proportions next generation will also be unchanged. The population is at equilibrium. But this equilibrium is quite different from the one when heterozygotes are favored.

Consider what happens if the population starts, not with equal numbers of *A* and *a* genes, but with 60% *A* and 40% *a*. Now if the heterozygotes make a smaller contribution to the gene pool, there will be a net increase in the proportion of *A* genes and decrease of *a* in the reproductive gene pool. To see this clearly, note that removal of heterozygotes takes out an equal *number* of *A* and *a* genes, but removing an equal number means removing a larger *proportion* from the class with the smaller numbers. So selection operates in a peculiar way: whichever gene is the more frequent tends to increase and whichever is least frequent tends to decrease still further. This, if not countered by some other force, will continue until the initially rarer gene disappears completely from the population.

We see now that the equilibrium when $p_A = p_a$ is an unstable one. It is like a stone balanced on a mountain peak. It has no tendency to move either way, but if once it is given a slight push, it will roll down. Selection in favor of heterozygotes, to continue the analogy, is like a stone resting in at the bottom of a crater; if it is displaced, it rolls back where it was. In the real world there are always momentary forces—if nothing more, the effects of random fluctuations in gene frequency due to the finiteness of the population—that carry the gene frequency away from an unstable equilibrium. Therefore, we should not expect to find populations that are polymorphic for loci where the heterozygote is at a disadvantage.

Our discussion has assumed that the two homozygotes are equally fit. This may not be true, of course, but the principle still works: as long as the heterozygote is less fit than either homozygote there is no stable equilibrium and the population tends to fix one allele or the other, usually the one that was initially most frequent.

One example of selection against a heterozygote is not a single gene, but a translocation. As you learned in Chapter 15, a translocation heterozygote is at a disadvantage compared to either homozygous gene arrangement because of the production of unbalanced gametes. Therefore you would not expect populations to be polymorphic for translocations. Whenever this is found, there is always some special, often tricky, explanation. New translocations are usually eliminated quickly from the population.

A second example is maternal-fetal incompatibility. When a fetus develops hemolytic disease because of Rh incompatibility it is necessarily heterozygous. If we designate the gene producing the antigen by *R*, the mother is necessarily *rr* or she would not be able to produce the antibody. The fetus necessarily has at least one *R* gene or its red blood cells would not have the antigen to react with the antibody produced by the mother. Of course, if the mother is *rr*, the Rh+ fetus must then be a heterozygote *Rr*. So every fetus destroyed by Rh incompatibility is heterozygous and there is selection against heterozygotes, just as we have been discussing.

Under these circumstances any population should tend to become homozygous for whichever alleles at the Rh locus were initially most common. This is true for most of the world's population. The Rh− gene is almost missing in Asiatics, American Indians, and Australoids.It is rare in most African populations, but has a frequency about .4 in populations derived from Western Europe. How do we explain the presence of both Rh+ and Rh− genes when our theory tells us that the rarer one should disappear? The most likely explanation is that the European population is of mixed origin. It is known that in the past there have been migrants to Europe from the East. These would have brought Rh+ genes. If the original European inhabitants had been Rh−, the mixed origin would be explained. A few thousand years have not been long enough for the kind of selection that we have discussed to have completed its work. The time scale is reasonable, for hemolytic disease is not common and we can see from the fact that Rh phenotypes in European and American adults agree very well with Hardy-Weinberg predictions that the selection is not very intense. There is evidence for a high Rh− frequency in the original European population. The Basques, in Southern France and Northern Spain, have the highest Rh− allele frequency of any part of the world, and for linguistic, geographical, and other reasons anthropologists believe that these come closest of any contemporary group to being remnants of the early Europeans.

PROBLEMS

1. About 1 calf in 256 among purebred Holsteins is born with red spots rather than black. Red is recessive. If none of the red reproduce, what would be the proportion of red calves next generation?
2. Starting with equal numbers of red and black genes, how long would it require to bring the proportion of red cattle down to 1 in 100? How many generations are needed to reduce the frequency from 1 in 100 to 1 in 900?
3. In swine the heritability of 6 months weight is 0.3. If the herd average is 180 pounds, the dam was 190 and the sire 200, what would be expected in the offspring?
4. A common strain of bean, Stringless Green Refugee, is susceptible to the bean mosaic disease virus. Another variety, less

desirable otherwise, has a dominant factor for resistance. How would you go about transferring the resistance factor into the susceptible strain?

5. There is another variety of bean with a recessive resistance factor. Would it be harder to transfer this into the susceptible variety? How could this be done?

6. A dominant mutant gene has a selective disadvantage, s, of 0.1. If the mutation rate to this gene is 0.00001, what will be the approximate frequency of persons showing the dominant trait at equilibrium?

7. If the gene is recessive, what will be the frequency of persons showing the trait?

8. If you like calculus, you might want to know how the numbers in the table on page 185 were computed. The change in one generation can be given by $\Delta p = p' - p = sp(1-p)[1 - h - p(1 - 2h)]/w$. If the process is slow (i.e., s is small), then w is very close to 1 and we can replace Δp by dp/dt. Can you integrate this for $h = 0$, 1, and 1/2, to obtain t as a function of s and the initial and final values of p?

9. Suppose the haploid number of genes in the human is 10,000 and that the rate of lethal mutation is 10^{-6} per gene. If these mutants are completely recessive, what would be the equilibrium number for lethal gene per genome? From what you know about the effects of cousin marriages, are the assumptions of this problem realistic? How would you amend them to make them more realistic?

10. Would you expect to find a large number of loci with as strong selection against homozygotes as is found with hemoglobin A and S?

11. Suppose, as in the example on page 188, that the frequency of the hemoglobin S allele is 0.1. Suppose malaria were eradicated from this area but that sickle cell homozygotes do not reproduce. How many generations would be required for the incidence of sickle cell anemia to decrease to 1/900? Approximately how many years would this be?

23

QUANTITATIVE INHERITANCE

QUANTITATIVE TRAITS

The gene-determined characters so far discussed have all been clear-cut qualitative traits, such as red and white eyes in flies, black and brown color in rabbits, smooth and rough hair in guinea pigs, and normal and abnormal hemoglobin in humans. There are also a whole series of traits, including many of great economic importance, in which the differences are in degree only. Examples of such characters are height, weight, milk production in dairy cattle, body conformation in beef cattle, yield in corn, and intelligence in man. These characters are all distinguished by the fact that many genes are involved and that usually the environment plays a large part in determining the phenotype. The inheritance of such characters is called **multiple factor** or **polygenic** inheritance.

A SIMPLE EXAMPLE

One example of this type of inheritance that lends itself to a simple hypothesis is found in crosses involving different seed colors in wheat. When a very dark red variety is crossed with a white type, the F_1 offspring are all intermediate in color. Then when these are self-fertilized, the F_2 show a variety of colors ranging from white to very dark red. Among the F_2 approximately 1/16 of the grains are as dark as the original very dark red parent and about 1/16 are as light as the white parent. This suggests the following hypothesis as to relation between genotype and phenotype:

$A'/A'\ B'/B'$	very dark red
$A'/A'\ B'/B,\ A'/A\ B'/B'$	dark red
$A'/A'\ B/B,\ A'/A\ B'/B,\ A/A\ B'/B'$	medium red
$A'/A\ B/B,\ A/A\ B'/B$	light red
$A/A\ B/B$	white

Apparently each of the primed letter genes contributes a small amount of red pigment, and there is no dominance.

The proportions in which these types would appear in the F_2 generation can be worked out by a Punnett square or other method and turn out to be

1/16 very dark red	(4 primes)
4/16 dark red	(3 primes)
6/16 medium red	(2 primes)
4/16 light red	(1 prime)
1/16 white	(0 primes)

Notice that these five proportions are exactly the same as the probabilities of obtaining four, three, two, one, and no boys in a family of four children. In this case it is the probability of four, three, two, one, or no primed letters out of four.

THREE PAIRS OF FACTORS

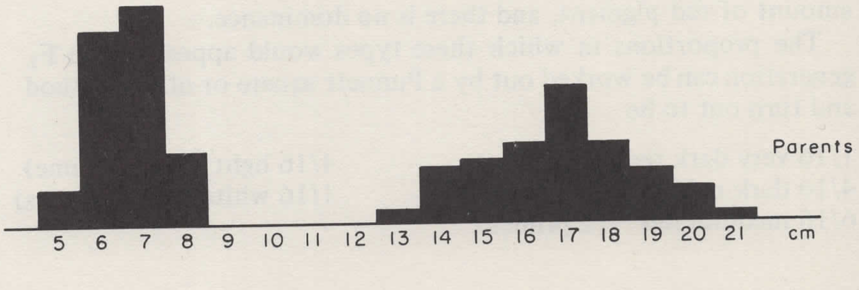
As a hypothetical example of the same type of inheritance with three pairs of factors, consider a plant where the genotype *AA BB CC* is 60 inches in height, and each addition of a prime adds one inch. The genotype *A'A'B'B CC* would then be 63 inches in height, and *A'A' B'B' C'C'* would be 66 inches. If we made a cross between *AA BB CC* and *A'A' B'B' C'C'* and then self-fertilized the F_1 generation the distribution of the F_2 would be

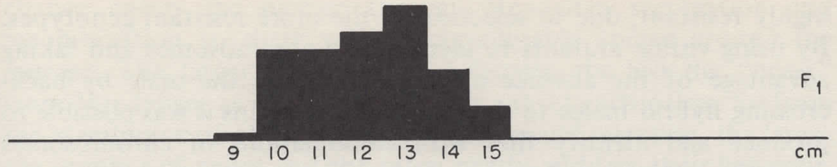
1/64	66 inches	(6 primes)
6/64	65 inches	(5 primes)
15/64	64 inches	(4 primes)
20/64	63 inches	(3 primes)
15/64	62 inches	(2 primes)
6/64	61 inches	(1 prime)
1/64	60 inches	(0 primes)

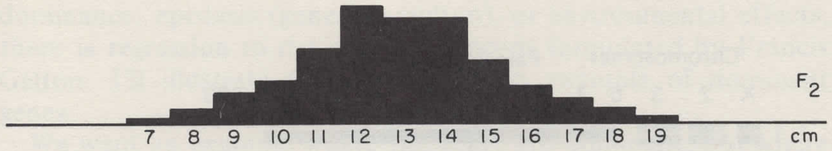
MORE THAN THREE FACTORS

In this case 1/64 would be as tall as the original tall parent. If there had been four pairs of factors, the proportion of the F_2 that would be as tall as the original tall parent would be $1/2^8$ or 1/256. Hence it is clear that if many genes are involved, a very large number of F_2 offspring would be needed to yield types as extreme as the original parents.

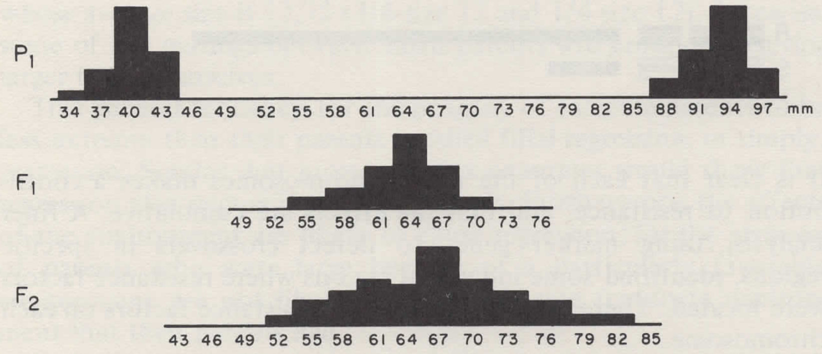
In most cases it is not possible to estimate the number of factors involved because of environmental fluctuations. But it is usually possible to see from the distribution of the parents, the F_1, and the F_2 whether there are many or few factors involved. Examples of these two types of situations are shown below. The first data show ear length in corn, measured in centimeters. In this case there are relatively few genes involved, perhaps four pairs or fewer. Notice that the average of the F_1 is somewhat nearer the long than the short variety length, indicating that there is some dominance of the factors for long length. Notice that there is a great deal of variability even in the parents and F_1.







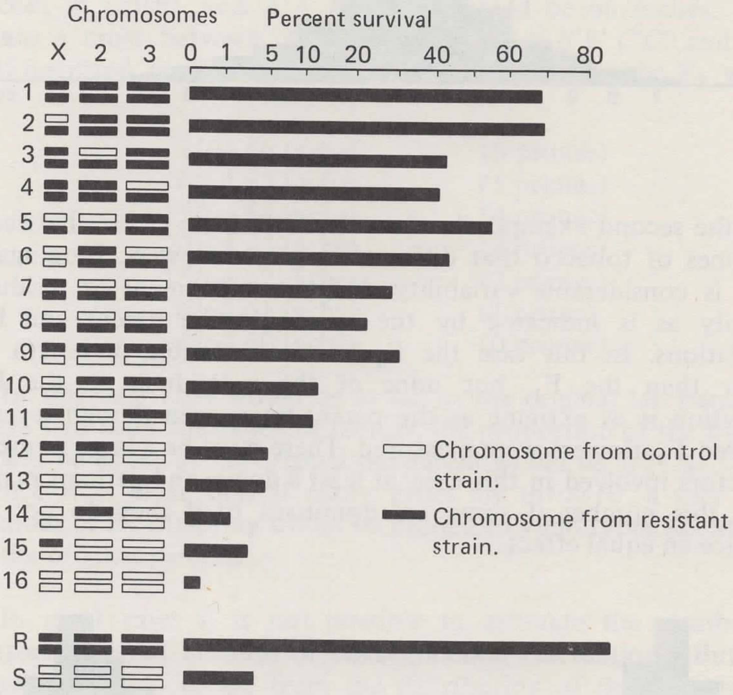
In the second example the data are taken from crosses between two lines of tobacco that differ in length of flowers. Here again there is considerable variability due to environment or residual heredity as is indicated by the spread in the parent and F_1 populations. In this case the F_2 is more variable in length of flower than the F_1 but none of the individuals in the F_2 generation is as extreme as the parent lines, even though several hundred F_2 flowers were measured. There must be a large number of factors involved in this case, at least a dozen and perhaps many times this number if some are dominant or if they do not all produce an equal effect.



LOCATING QUANTITATIVE FACTORS BY CHROMOSOME ANALYSIS

By using marker genes it has sometimes been possible to dissect the quantitative trait and find the chromosomal location of genes having quantitative effects. Usually this is hard to do because of the smallness of the individual gene effects. One approach is to study one chromosome at a time.

A good example is provided by polygenic DDT resistance in Drosophila. The resistant strain was developed by putting increasing amounts of DDT into a cage where a large population of flies was growing. After many generations the strain had become highly resistant, due to selection of the more resistant genotypes. By using visible mutants to identify the chromosomes and taking advantage of the absence of crossing-over in the male by back-crossing hybrid males to the two parental strains it was possible to produce and identify flies with combinations of chromosomes from the resistant and normal strains. The results are shown in the graph below.



It is clear that each of the major chromosomes makes a contribution to resistance, and that the effects are cumulative. A finer analysis, using marker genes to detect crossovers in specific regions, identified some individual regions where resistance factors were located. There appear to be several resistance factors on each chromosome.

REGRESSION TO THE MEAN

The models of gene action that have been considered in this chapter have assumed additive genes with no dominance. Such a system has one consistent property: the expected average of the progeny is always the same as the average of the parents. In the wheat example at the beginning of this chapter, the progeny of dark red (say *A'/A' B'/B*) and light red (say *A'/A B/B*) will average medium red (25% dark red, 50% medium red, and 25% light red) just like the parents.

As long as there is no dominance, no epistasis, and no environmental effect this distribution pattern will be true. Of course, chance enters the picture through Mendelian segregation and recombination so there will be considerable spread around the average. The larger the number of factors, the less the "noise" from Mendelism as the various effects cancel each other. So from the standpoint of efficiency of prediction and selection, the larger the number of genes and the more nearly additive their behavior, the happier the breeder is.

The real world doesn't provide such ideal conditions. If there is dominance, epistasis (gene interaction), or environmental effects, there is **regression to the mean**, a concept formulated by Francis Galton. I'll illustrate regression with an example of dominant genes.

We want an example where the genes are dominant, but where the behavior of F_1 and F_2 resemble superficially the additive systems that we have been considering. We can do this by making half the genes dominant in one direction (say, toward smallness) and half dominant the other way. The simplest example involves two pairs of genes. Assume that A and b tend to increase size, while a and B decrease it.

Genotypes	Size (in arbitrary units)
$A/-\ b/b$	13
$A/-\ B/-$ or $a/a\ b/b$	12
$a/a\ B/-$	11

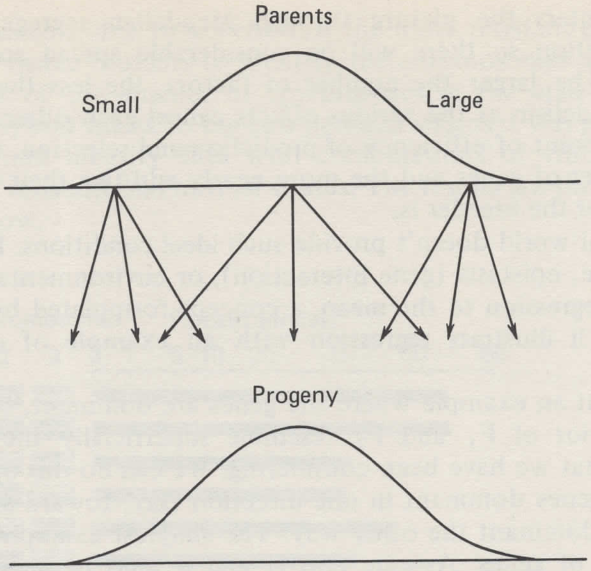
Now, if various hypothetical crosses are made, the average of the progeny will sometimes be the same as the average of the parents. But the progeny of some of the crosses will not be. For example, two parents of genotype $A/a\ bb$ of size 13 will produce progeny whose average size is 12.75 (3/4 size 13 and 1/4 size 12). Likewise, some of the matings between small parents will produce offspring larger than themselves.

This general tendency for the progeny of extreme parents to be less extreme than their parents is called **filial regression**, or simply, regression. Similar, but more complex examples would show that regression also occurs if there is epistasis. Furthermore, the effects of the environment are likely to cause regression; for the progeny of parents who were large because of a particularly favorable environment are not likely to have the same fortunate environment that their parents had.

Each generation the large parents have progeny, some of which are larger than themselves but mostly smaller. Likewise, the small parents produce progeny that average slightly larger than themselves. Hence the progeny of parents that are at either end of the scale tend to regress toward the population average. On the other hand, a few of the progeny of average parents are extreme so that the total distribution of progeny is like that of the parents.

The relationship can be diagrammed as shown on the following page.

This means that heredity isn't as predictable as it would be if all the genes acted additively with no dominance. The phenomenon


Parents
Small
Large
Progeny

of regression, plus the random variations imposed by Mendelian segregation and recombination, make heredity less precise (though perhaps more interesting) than it otherwise would be.

To correct for the effects of regression, animal and plant breeders use a quantity called **heritability**.

SELECTION OF QUANTITATIVE CHARACTERS

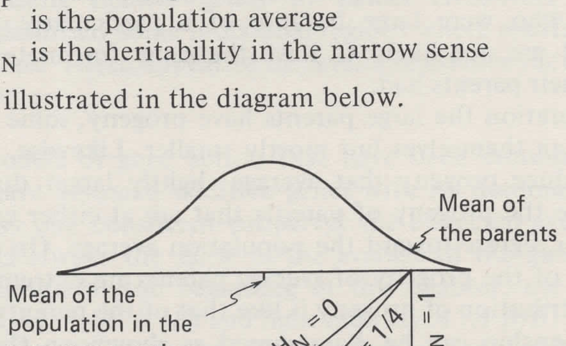
It has been observed that the fraction by which the progeny of extreme parents regresses toward the population average is approximately constant, regardless of the extent of the deviation from the mean. This makes possible a very simple formula for predicting the offspring from measurements on the parents. Letting Y stand for yield (or performance, or whatever quantitative trait is of interest)

$$Y_O = \bar{Y} + H_N(Y_P - \bar{Y}) \tag{23-1}$$

where

Y_O is the predicted average value of the offspring
Y_P is the average of the two parents
$\bar{Y}$ is the population average
H_N is the heritability in the narrow sense

This is illustrated in the diagram below.


Mean of the parents
Mean of the population in the parent generation
$H_N = 0$
$H_N = 1/4$
$H_N = 1$
Mean of the progeny

(H_N is used to designate *heritability in the narrow sense*, to distinguish it from a different quantity, *heritability in the broad sense*. Both will be discussed in the next chapter.)

For example, assume that the heritability of 160-day weight in swine is .30. If a breeder selects from a population averaging 180 pounds a group averaging 210 pounds to use as parents, the offspring would be expected to average 180 + .30(210 − 180) = 189 pounds.

As long as heritability figures are known, predicting the effects of a selection program is quite feasible. There are always chance fluctuations, especially if only a few individuals are involved, but these cancel out with large numbers. The real difficulty is that the heritability is hard to determine. It is usually found from the correlation between various relatives and requires measurements on hundreds of individuals to be reliable. Furthermore, the figure is relevant only to the population in which it was determined. Heritability data on weight in hogs would have no relationship to that of milk production in cattle, and would even be different for another hog population with a different amount of inbreeding or a different environment. Nevertheless, heritability data are of great utility to the practicing breeder, and here are a few typical figures:

Table 23-1

Animal	Character	Heritability, H_N
Cattle	Amount of white spotting	0.95
Chickens	Egg production	0.20
Swine	Birth weight	0.06
Swine	Rate of growth	0.30

SELECTION IN FARM ANIMALS

One has only to look at the variety of forms of farm animals to see what selection can do. The great differences between beef and dairy cattle are the result of only a few hundred years of selection. Animal breeders have not until this century been able to make use of knowledge of Mendelian heredity, but they knew that superior offspring came from superior parents.

The present-day breeder has more knowledge at his disposal. There are several ways he can make his selection more effective. One is to concentrate on important characters. If he is interested in milk production he can make sure that he is using this as a criterion and not some less important character in choosing his breeding stock. A second thing he can do is to make use of pedigree information. Other things being equal, a bull that comes from a family of high producers has a better chance of transmitting genes for high production than one who doesn't. Thirdly, he can make use of **progeny testing**. He can select not on the basis of the merit of the individual but on the merit of his offspring. This is particularly important in a character like milk or egg production where only one sex can be judged. Progeny test selection is now used extensively in dairy cattle programs where artificial insemination is practiced.

In order to make the best progress by selection the animal breeder can use a **selection index**, a device by which a particular animal's phenotype, ancestry, and available progeny are combined with other relevant information to provide an estimate of his worth as a breeder. Then those animals with the highest indices are selected as breeding stock.

SELECTION IN ECONOMIC PLANTS

One of the most important applications of selection has been in developing disease-resistant varieties of plants. It is fortunate that resistance is often due to a single gene which may easily be selected for or transferred from one strain to another. Almost every variety of cultivated plants has been improved by incorporating into its germ plasm one or more genes for resistance to some important disease.

Also by continued selection there has been great improvement in yields. Selection for yield in corn is an illustration of the use of successively better methods. At first mass selection was practiced. Seed was selected from the best ears each generation. Next a form of progeny test selection, the "ear to row" method, was used. The kernels from each plant were planted in a single row and the selection was made on the basis of the production of the entire row. In this way the genes from those plants capable of transmitting high yield to their descendants were selected. Finally a still better method, applicable at least in corn, has been used—that of producing hybrids from selected inbred lines, thus combining selection with heterosis.

SELECTION AND EUGENICS

Most people hope that the world of the future will contain people who are healthier, more intelligent, and happier than the present population. Or at least they hope that the population will not deteriorate in any of these or other desirable qualities. Since most characters are due to combined effects of genes and environment, it is likely that some change could be brought about by changing either factor. The improvement of the population by genetic change has been called **eugenics**; changing the population by improving the environment is **euthenics**.

In the early days of genetics, eugenics became a very popular cause taken up by many geneticists and many more nongeneticists. Judging from the results of selective breeding in domestic animals, many eugenists assumed that by selection the human population could be changed rapidly into any desired type and that mental and physical disease could be greatly reduced. Now geneticists view the question as one which is much more complex. Selection is a slow process in any animal or plant with a long life span and a small number of offspring per family. As we have seen, it is particularly slow against recessive genes.

On the other hand, genetic knowledge increases every year. For many recessive diseases the carrier state can now be detected, so that much more efficient selection could be practiced if this were desired. Eventually the techniques of molecular biology and of microbial genetics will become applicable to man. Even now the possibility of selection by artificial insemination exists and has been strongly advocated by some, notably H. J. Muller.

There has been considerable discussion of the possible consequences of medical progress and a higher standard of living. Doubtless a number of harmful genes that in a more rigorous society would have been eliminated by natural selection now are transmitted to future generations. Such effects are not likely to be sudden nor catastrophic, but they do point up the fact that man will some day have to exercise some conscious influence on his genetic future.

Clearly there is no general agreement now as to the extent to which man should try to influence his genetic future. How much society should try to influence the reproductive practices of this generation in order to have a better supply of genes in the future is an involved question and geneticists alone cannot answer it.

PROBLEMS

According to Davenport, skin color in Negro-white marriages is inherited as if there were two pairs of genes without dominance, exactly like the genes that determine seed color in wheat. The five color classes may be called black, dark, medium, light and white. This hypothesis is almost certainly an over-simplification, and therefore wrong. However, it is a crude approximation to the truth, and may be used as a basis for answering the first four following questions.

1. What is the darkest colored child that could result from two parents, one light and one white?
2. What is the darkest colored child that could result from two parents, both light?
3. What is the lightest colored child that could result from two parents, one black and one dark?
4. Sometimes two intermediate parents produce children all of intermediate color; other similar appearing parents produce children ranging from very light to very dark. Suggest an explanation.
5. In the two-factor dominance model on page 197, what would be the expected progeny from a cross of *A/A b/b* x *a/a BB*? If these progeny are mated among themselves, what ratios would be expected in the F_2?
6. A cross between a six-foot and a four-foot plant produced progeny that were uniformly five feet. In the F_2 there were 1/256 of the plants four feet and 1/256 six feet. Assuming that the factors involved are equal and additive and without dominance, by how many pairs of factors do the two parents differ?

7. A six-foot x four-foot cross gives, as before, uniform five-foot progeny. But this time 9/256 of the F_2 are six foot and 9/256 four footers. Explain these results, assuming dominant factors with an equal number tending toward tallness and shortness.
8. If we class mentally retarded children into severely retarded and mildly retarded, one observes that the average IQ of the sibs of the severely retarded children is higher than the average IQ of the sibs of the mildly retarded. Suggest an explanation for these paradoxical findings.

24

STATISTICAL ANALYSIS OF QUANTITATIVE CHARACTERS

Since quantitative characters are usually polygenic and strongly influenced by environment, it is ordinarily impossible to analyze the effects of individual genes. The alternative is to use statistical procedures derived from our knowledge of gene behavior and study the properties of aggregates of genes without trying to determine how the individual genes are acting. The methods were developed mainly by R. A. Fisher and Sewall Wright, and depend on correlations between relatives and analysis of variance.

MEASURES OF CENTRAL TENDENCY

We frequently describe an unwieldy mass of data with one or two summarizing constants. We usually want some indication of the center of the distribution and some measurement of how much variability there is around the center. Sometimes other measures are given, such as measures of the degree of asymmetry.

The usual measure of the location of the center is the arithmetic mean. If we have a series of measurements on N individuals, the individual measurements being $x_1, x_2, x_3, \ldots\ldots x_N$, the mean is defined as

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N} = \frac{\Sigma x}{N} \qquad (24\text{-}1)$$

where the Greek sigma (Σ) stands for the sum.

Sometimes the **median** is used. This is the center measurement; that is, it is the measurement that divides the group into two halves, with half of the values above it and half below.

For some purposes the geometric mean is convenient. This is the N^{th} root of the product of the N measurements. Usually it is obtained as follows: we take the logarithm of each measurement, then obtain the arithmetic mean of the logs, then take the antilog of the arithmetic mean. This works because multiplying quantities is equivalent to adding their logs. Such a procedure is very useful when the genes appear to act in a multiplying manner rather than in an additive manner.

THE MEASUREMENT OF VARIATION

The conventional measurement of variability is the **variance**. The variance of a population is best estimated from a sample by the formula

$$V = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_N - \overline{x})^2}{N - 1}$$

$$= \frac{\Sigma\,(x - \overline{x})^2}{N - 1} \qquad (24\text{-}2)$$

where $\overline{x}$ is the mean of the sample as defined in the previous section and N is the number of individual measurements.

The variance is thus the average of the squared deviation from the mean. Clearly it will be larger for a variable population than for a constant one. The reason for dividing by N − 1 instead of N is that we are using the sample to estimate the variance of the population from which it was drawn. The sample variance tends to underestimate the population variance and the underestimate is exactly corrected by dividing by the number of degrees of freedom, N − 1, instead of the number of measurements. If we are interested in the sample itself instead of its parent population, we divide by N.

The variance has the disadvantage that it is measured in squared units. For example, if we are studying a group of measurements of height in centimeters, the variance is in square centimeters. It may even appear to be quite meaningless as when the variance of corn yield is measured in square bushels.

One way of overcoming this dimensional disadvantage of the variance is to take its square root. This is called the **standard deviation** and is abbreviated s or σ. (Frequently one uses Greek letters to stand for the unknown true values in the population and Roman equivalents for their estimates derived from a sample.) Thus, the estimated standard deviation of x is given by

$$s_x^2 = V_x \qquad (24\text{-}3)$$

If the data are normally distributed (that is, if they conform to the bell-shaped normal or Gaussian curve) we can say, as shown in Appendix II, that about 67% of the measurements lie within one standard deviation of the mean and that 96% are within two standard deviations.

Appendix III shows how the standard deviation may be used to make significance tests for measurement data, analogous to the χ^2 test for enumeration data.

Unfortunately, the standard deviation does not have the property of additivity which is usually more important in genetic analysis than dimensional simplicity, so for most genetic purposes the variance is used as the measure of variation. The additivity and subdivisibility of the variance are illustrated in the next section.

SUBDIVISION OF VARIANCE

If the variance of a population of animals or plants is estimated by the formula given above, the value so obtained is the total or phenotypic variance, designated by V_T.

One may then ask the question: what would the variance be if the environment had been held constant? The answer obtained by such an experiment would be the **genotypic variance**, the variance due to genetic differences alone. This is designated by V_H. Likewise, if the genotypes could be held constant, the remaining variance would be due to differences in environmental factors, and this could be designated as the environmental variance, V_E. Because of the additivity of the variance, the total variance must be the sum of the genotypic and environmental variance (provided these are independent), or in symbols

$$V_T = V_H + V_E \tag{24-4}$$

It is usually impossible to hold the environment constant, but it is possible to hold the genotype constant in animals that have identical twins or in plants that can be propagated vegetatively. Hence, it is possible, in principle, to measure the proportion of the variance that is due to hereditary differences and that due to environmental differences, though in actual practice it may be very difficult.

The genotypic variance itself may be further subdivided. One might ask, in the same manner as above, how much variance would remain if the genes were without dominance and epistasis. One can divide the hereditary variance into a fraction due to additive gene effects (V_G), a fraction due to dominance (V_D), and a fraction due to epistasis (interactions, V_I). Thus

$$V_H = V_G + V_D + V_I \text{ or} \tag{24-5}$$

$$V_T = V_G + V_D + V_I + V_E \tag{24-6}$$

I do not intend to imply that these quantities have much relation to what individual genes are actually doing, but on the average the cumulative effect of all the genes can be described in this fashion. And this form of description, artificial though it is in some respects, is the most useful procedure yet devised.

COVARIANCE AND CORRELATION

The covariance, C_{xy}, of two variables, x and y, is analogous to the variance. We consider that the measurements are somehow grouped in pairs. For example, x_1 and y_1 may be measurements of the heights of two brothers, x_2 and y_2 the heights of a second pair of brothers, and so on. The covariance is estimated from the sample by the formula.

$$Cxy = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \ldots + (x_N - \bar{x})(y_N - \bar{y})}{N - 1}$$

$$= \frac{\Sigma(x - \bar{x})(y - \bar{y})}{N - 1} \tag{24-7}$$

Notice that if the two members of a pair are totally independent of each other the covariance on the average will be zero, since the two members of a pair will deviate from their averages in opposite directions just as often as in the same direction, and there will be just as many minus terms as plus terms. Notice, too, that the covariance (unlike the variance) can be negative; it will be if the two members of a pair tend to deviate in opposite directions from their averages. Covariances, like variances, are additive.

The covariance is often divided by the geometric mean of the two variances to give the **coefficient of correlation,** r_{xy}. Thus,

$$r_{xy} = \frac{C_{xy}}{\sqrt{V_x V_y}} = \frac{\Sigma (x - \overline{x})(y - \overline{y})}{\sqrt{\Sigma (x - \overline{x})^2 \; \Sigma (y - \overline{y})^2}} \tag{24-8}$$

The value of r ranges from +1 (complete association of x and y) through 0 (no association) to −1 (complete negative association). Notice that this is reasonable, for if $x_1 = y_1$, $x_2 = y_2$, etc., then $r_{xy} = 1$; whereas if $x_1 - \overline{x} = -(y_1 - y)$ and so on, $r_{xy} = -1$.

INTERACTION AND COVARIANCE

The simple formula relating the genotypic and environmental to the total variance Equation 24-4 depends on the genotypic and environmental factors being independent. A more general formula is

$$\underset{\substack{\text{Total}\\ \text{(Phenotype)}}}{V_T} = \underset{\substack{\text{Heredity}\\ \text{(Genotype)}}}{V_H} + \underset{\text{Environment}}{V_E} + \underset{\text{Interaction}}{V_{HE}} + \underset{\text{Covarience}}{2C_{HE}} + \underset{\text{Error}}{V_e} \tag{24-9}$$

Where,

V_T is the measured variance of the trait being considered in the population.

V_H is the variance that is attributable to differences in genotype. It would be zero if all the individuals in the population were of the same genotype.

V_E is the variance due to environmental differences. It would disappear if it were possible to raise all the individuals in the same environment.

V_{HE} is the variance caused by nonadditivity of genetic and environmental effects. If a change in the environment has a different effect on different genotypes, this creates such an interaction.

C_{HE} is the covariance of heredity and environment. It arises if the genetic and environmental factors are not randomized. For example, if a farmer places his cows with the best genotypes in the best pastures, this would produce a covariance.

V_e is the variance due to errors of measurement or other uncontrollable random effects.

We might, for example, grow a number of strains of corn in different fields in different locations. We assume that all the plants of a strain are genetically identical and that the strain can be reproduced (as if each strain is produced by crossing two highly inbred lines). Later, we shall divide V_H into subcomponents based on the manner of gene action; but this requires that we take Mendelian segregation and recombination into account, along with the mating system. So, for the present, we consider that the strain can be propagated asexually or reconstituted by crossing homozygous lines.

If V_H is large, this is an indication that we can improve the performance by choosing the better strains and planning a breeding experiment to produce still better ones. If V_E is large, we look for ways to improve the environment. This isn't always easy, for we may not know what the relevant environmental factors are. This depends on knowledge of the subject—the experimenter may notice that the fields that did best are those of a particular soil type or with some other identifiable property.

Interaction variance is often troublesome to deal with. If it is small, it is often simply ignored. If it is systematic—for example, if improved environment causes the largest increase in yield in plants with the best genotypes—then we will make more progress by changing both the genotype and the environment than by changing only one. If it is not systematic, the breeder may need to develop special strains for different environments.

This has been done with corn, where there are specific hybrid strains for particular growing seasons and soil types.

In planned experiments, covariance is not a problem, for strains and environments are randomized. But in the study of natural populations, and in particular in the study of human populations, this is a factor to be reckoned with. To some extent the genotype creates the environment, as when a child with a talent for music is sent to a school that specializes in music.

Errors of measurement are usually not a serious problem, but a conspicuous exception to this statement is human behavioral traits. Intellectual, and especially emotional, traits present serious problems if one is trying to find reproducible and valid measurements.

TWO KINDS OF HERITABILITY

Heritability is defined in two ways and the two have different purposes.

a. **Heritability in the broad sense,** H_B, is an attempt to measure the importance of genetic factors for the trait under consideration. It measures the extent to which individual differences in the population are due to differences in genotype. It is useful in determining the relative influence of genetic and environmental factors in causing individual differences.
b. **Heritability in the narrow sense,** H_N, is a measure of the extent to which phenotypic differences are transmitted from parents to their progeny. It is used to predict the amount of improvement in a population that can be brought about by selective breeding.

Some confusion has been introduced by the fact that geneticists typically use heritability in the narrow sense whereas psychologists use it in the broad sense. This is understandable, for geneticists are usually concerned with predicting the effect of selection in programs of plant and animal breeding, whereas psychologists are usually more interested in assessing the extent to which genetic and environmental factors are responsible for individual differences.

We shall first consider heritability in the broad sense.

HERITABILITY IN THE BROAD SENSE

The broad sense heritability, H_B is defined as

$$H_B = \frac{V_H}{V_T} \tag{24-10}$$

and thus is the fraction of the total (phenotypic) variance that is attributable to genetic differences.

This type of heritability is measured most directly by placing individuals of identical genotypes in random environments. With many kinds of plants identical genotypes can be obtained by vegetative propagation. In higher animals, including ourselves, we can take advantage of one-egg twins.

If we can ignore the genotype environment interaction ($V_{HE} = 0$), then the heritability is measured by the correlation of one-egg twins (or other types of identical genotypes) grown in randomized environments. This is easily shown by noting that the phenotypic value of an individual is $t = h + e$, where t, h, and e are all measured as deviations from the mean. The quantities h and e represent genotypic and environmental deviations. A second individual of the same genotype has the deviation $t' = h + e'$, since the environments are different while the genotypic value stays the same. The covariance of identical twins is

$$\begin{aligned} C_{TT'} &= \frac{\Sigma(h + e)\ (h + e')}{N - 1} \\ &= \frac{\Sigma h^2}{N-1} + \frac{\Sigma he}{N-1} + \frac{\Sigma he'}{N-1} + \frac{\Sigma ee'}{N-1} \\ &= V_H + 2C_{HE} + C_{EE'} \end{aligned} \tag{24-11}$$

since it is reasonable to think that $C_{HE} = C_{HE'}$. If the environments are independent of the genotypes and independent of each other, then the expected values of the last two terms are zero. So the phenotypic covariance is the genotypic variance.

The correlation coefficient is the ratio of the covariance to the geometric mean of the two phenotypic variances. Since the latter, V_T, is the same for the two twins, the correlation is V_H/V_T, which is the definition of the broad sense heritability. So we can say

$$H_B = r_{TT'} \tag{24-12}$$

provided the twins are reared in independent environments and the genotype-environment interaction can be neglected.

The broad sense heritability of milk and butterfat yield of cattle, as measured by this method, is about 0.9. The heritability of size is about 0.8, suggesting that in the environments where the twin studies were done (presumably representative of good farming practice), genetic differences account for much more of the variance than environmental differences.

Human identical twins reared in different homes are very similar in height ($H_B = .95$), weight (.9), and IQ (about .8). These would suggest high degrees of genetic determination of the three traits, but there is always the question whether human identical twins adopted into different homes are actually put into homes that are equivalent to a random sample. If the homes are more similar environmentally, then the heritability is overestimated. Prenatal effects and any influence of the common environment before the twins were separated would also inflate the estimate. Finally, the number of separated twins is small and there are errors of measurement, especially for the IQ scores. So the value of the heritability for IQ remains in doubt.

The example of height illustrates what heritability tells us and what it does not. It tells us that in the existing circumstances differences in genotypic differences are more important than environmental differences in determining height. But we know that extremes of environment can have a large influence. We know that the height of the Japanese has increased a great deal since World War II (as well as the height of other racial groups) and that this gain in stature cannot be of genetic origin. Thus heritability cannot predict the effect of a uniform environmental factor applied to the whole population.

THE THEORETICAL CORRELATIONS BETWEEN RELATIVES

We learned before that variances and covariances are additive if the component measurements are independent. That is to say, the variance of the sum of a series of independent measurements is the sum of the individual variances, and likewise for the covariance. Two related individuals are identical for a certain fraction of their genes. The remaining genes are independent. The covariance of the contributions of the independent genes is zero. The covariance of the contributions from the identical genes will be the same as the variance. Thus, on the average, the covariance will be ρ times the variance, where ρ is the proportion of genes that are identical in the two individuals. Thus the theoretical correlation coefficient is ρ, the fraction of genes in common.

The fraction of common genes is easily determined. A parent and offspring have 1/2 their genes in common. Likewise, half-sibs have 1/4. In general the fraction of common genes is twice the coefficient of kinship, if the population is not inbred.

HERITABILITY IN THE NARROW SENSE

The narrow sense heritability, H_N, is defined as

$$H_N = \frac{V_G}{V_T} \qquad (24\text{-}13)$$

This measures the fraction of the phenotypic variance that is due to the additive component of each gene, and thus is a measure of the average contribution from parent to progeny.

Roughly, the narrow sense heritability is measured by the ratio of the observed correlation between relatives to the theoretical value.

$$H_N = \frac{\text{Observed correlation between relatives}}{\text{Theoretical correlation between these relatives}} \quad \text{(rough definition)} \qquad (24\text{-}14)$$

The theoretical correlation is the proportion of genes shared by the relatives. For parent and offspring the fraction of shared genes (ignoring X- and Y-linked loci) is 1/2. So, one measure of H_N is the observed correlation between parent and offspring divided by 1/2. There is a disadvantage, however, because it is hard to arrange things so that a parent and offspring have independent environments. The relationship that is preferred by animal breeders is half-sibs with a common father and unrelated mothers, with the offspring reared in independent environments. (The other kind of half-sibs are less desirable, for there are common maternal influences, prenatal and postnatal, that might be confused with genetic factors.) Since half-sibs share 1/4 of their genes, the heritability is four times the half-sib correlation.

The proportion of shared genes, when the population is mating at random, is twice the coefficient of kinship. For example, with half-sibs the coefficient of kinship is 1/8, therefore the narrow sense heritability is four times the observed correlation of half-sibs; and similarly for other degrees of relationship.

Note: This formula does not work for full sibs. The reason is that sibs can share both of their alleles at a given locus, since they are connected by two paths of ancestry. The other relatives mentioned above share only one allele at a locus. If two alleles can be shared, then dominance makes a contribution to the correlation. (The easiest way to visualize this is to think of rare recessive genes; two sibs can both be homozygous, whereas it is extremely rare for both parent and offspring to be homozygous for such a gene.) Thus part of the full sib correlation should be attributed to the dominance variance. In fact the amount by which the sib correlation exceeds the parent-offspring correlation (or twice the half-sib correlation) can be used to measure the dominance variance, if the environmental effects can be properly randomized.

This formula does not take care of epistasis. In fact, there is no satisfactory way in practice. Fortunately, the epistatic terms enter the correlations accompanied by very small coefficients, so they do not contribute much to the correlation. Empirically, prediction formulas that ignore epistasis work pretty well. Whether this means that epistatic interactions for quantitative traits are small, or whether they are in opposite directions and cancel each other,

that is due
measure of

ɔy the ratio
theoretical

latives

(24-14)

s shared by
hared genes
re of H_N is
divided by
l to arrange
nt environ-
breeders is
rs, with the
her kind of
ernal influ-
vith genetic
heritability

n is mating
ample, with
the narrow
on of half-

ne reason is
, since they
er relatives
alleles can
the correla-
re recessive
extremely
for such a
tributed to
ich the sib
(or twice
the domi-
e properly

there is no
terms enter
ts, so they
prediction
hether this
s are small,
each other,

---

or whether the measures are so insensitive as to miss such effects, is not clear.

Some examples of the value of narrow sense heritabilities and their use in predicting the effects of selection were given in the previous chapter.

PROBLEMS

1. In his early studies of guinea-pig coat color, Wright found that the total variance in a randomly mating strain was .573. The variance in an inbred strain, which may be assumed to be homozygous, was .340. The heritability, determined from correlation between relatives in the randomly mating population, was .38. Fill in the following table.

	Variance component	Component as percent of total
Additive genetic variance		38%
Dominance and epistasis		
Environment	.340	
Total	.573	100%

2. The heritability within the inbred population was not significantly different from zero in Wright's study. Is this in accord with genetic theory?

3. In swine, the estimate of narrow heritability of birth weight is 6%. Which character could more quickly be changed by coat color in guinea pigs or birth weight in pigs?

4. The correlation between half-sibs for body weight at 180 days in swine is .075. What is the heritability?

5. Here are some hypothetical data on heights in a plant population. The first is for a randomly mating population in a standard environment. The second group of plants is grown in an absolutely constant environment (assume that this is possible). The third group is from a population that is genetically identical, being propagated from one plant by vegetative means.

Random mating ordinary environment	Random mating constant env.	Identical genotype ordinary env.
105	103	100
102	102	104
96	99	98
98	101	100
99	98	98
100	98	102
104	102	99
102	104	99
101	96	101
95	97	99
98	100	100

Compute the mean and variance for each column. What is the total variance, V_T? What is V_H? What is V_E? Can the heritability be determined from these data? Are the data internally consistent, i.e., are the variances properly additive?

6. A formula for the variance that is more convenient for computation is

$$V = \frac{\Sigma x^2 - N\bar{x}^2}{N-1}$$

If you like algebra, show that this is equal to the formula given at the beginning of this chapter.

7. Show that if each y is a constant multiple of the corresponding x, their correlation is one. Hint: Let $y_1 = kx_1$, $y_2 = kx_2$, etc., where k is a constant.

GENETICS AND EVOLUTION

25

HISTORICAL IDEAS OF EVOLUTION

I shall not go into the evidences for and against the occurrence of evolution, but will take for granted that it has occurred and is still occurring. The contribution of genetics to the study of evolution has been to explain how evolution occurs, and it is that with which we shall be concerned.

If we ignore all theories implying purposeful evolution or requiring some supernatural force as being incapable of being tested by scientific methods, we have as the principal foundations for present concepts the observations and ideas of four men, Lamarck, Darwin, Weismann, and De Vries.

1. *Lamarck* (1744-1829). The Lamarckian idea is that acquired characters and the effects of use and disuse are inherited. An animal which had exercised a great deal and developed strong muscles would have offspring with stronger muscles. An organ that is not used will atrophy and in the next generation will be smaller. A man who spent a great part of his time outside and acquired a suntan would transmit this to his children.

   Much of the long-continued popularity of Lamarck's idea is due to its success in "explaining" the adaptation of plants and animals. Progressive evolution would be expected under this system of heredity. However, the results of numerous laboratory experiments in the past 50 years have shown that practically without exception acquired characters are not hereditary. There are numerous other objections, one being that it is hard to conceive of the physiological mechanisms by which the phenotype can influence the germ cells in a purposeful manner. Another difficulty with a Lamarckian inheritance is explaining how sterile classes in the social insects, such as worker bees, become adapted to their environment. It can hardly be expected that the queen can transmit characters acquired by the workers! For these and many other reasons the Lamarckian theory, attractive as it is, has been largely rejected.
2. *Darwin* (1809-1882). Darwin confined his attention to the variations that are inherited, irrespective of how they were acquired. He knew such hereditary variations existed although at that time the precise mechanism of their inheritance was not known.

Darwin's reasoning has been summarized by Julian Huxley in this way. He starts with three observable facts:

a. In virtually every species of plants and animals the offspring produced are more numerous than their parents.
b. Nevertheless most populations are relatively stable in size.
c. There is a great deal of variability in nature, and some of the variability is hereditary.

From the first two observations it follows that there is competition for survival. Because of the variability in the population some individuals have a better chance of surviving than others, and to the extent that variations making for greater survival are hereditary, these variations will tend to increase in proportion in future generations. This process is **natural selection.**

Darwin was not familiar with what we know of the mechanism of heredity. He had several ideas on the subject and in his later years accepted Lamarckian inheritance as at least part of the explanation. The principal addition of modern genetics to Darwin's theory has been the knowledge of the origin of variations by mutation and their mode of transmission by Mendelian heredity.

3. *Weismann* (1834-1914). The great contribution of Weismann was to distinguish clearly between the germ cells and the somatic tissues and to point out that effects on somatic cells do not ordinarily affect the germ cells and hence are not passed on to the next generation. Weismann's famous experiment of cutting off tails of mice for many generations only to find the descendants still born with tails did much to change the minds of biologists from the Lamarckian views prevailing at that time.
4. *De Vries* (1848-1935). Besides being one of the three discoverers of Mendel's work, De Vries is responsible for the idea of mutations as a force in evolution. It has subsequently turned out that the "mutations" with which De Vries worked were really segregations from complex chromosome rearrangements and that most mutations produce much smaller effects than he thought, but De Vries must be credited with being the first to point out that mutations provide the source of the variability on which natural selection acts.

BLENDING VS. PARTICULATE INHERITANCE

Previous to the general recognition of Mendel's laws, inheritance was thought to be a blending process. The mixing of germ plasms was considered analogous to the mixing of fluids, which could not again be separated into their component parts. For example, the offspring of a cross between a small and a large plant was thought to be intermediate (as is often the case) but the F_2 plants were considered to be just as uniform as the F_1. Now, thanks to Mendel, we know that hereditary characters are carried by particulate units. In the F_2 the genes are reassorted and the variability remains.

The evolutionary significance of these two ideas lies in the fact that with a blending theory of heredity the loss of variability each generation is so great that there would soon be no hereditary differences for natural selection to act on. In order to account for the existing variability there must be some source of new variations to replace those lost by hybridization. On the other hand, with a particulate theory of heredity the variability is conserved *and a very small mutation rate is sufficient to keep the population in a highly variable state—one in which evolution is possible.*

CONTEMPORARY EVOLUTION THEORY

The basic ideas of Darwin as refined by subsequent knowledge are now accepted by practically all biologists. Simply stated, the prevailing ideas are as follows:

1. Gene mutations (together with chromosome rearrangements) occur at predictable rates, but in unpredictable individuals quite by chance. These provide the basis of variation.
2. Through the processes of meiosis and fertilization these mutants are combined in various ways with other factors.
3. Genes that are responsible for characters that favor the survival of the individual tend to increase in relative frequency while harmful genes become less numerous.
4. This process of natural selection results in the species gradually changing and becoming successively better adapted to its environment, and over millions of generations is sufficient to account for evolution as it has occurred.

THE TIME SCALE

The remarkable adaptations that have occurred since the earliest organisms don't seem so inexplicable when we consider the long time scale involved. Here are some figures, based on the fossil record. Some of the values are quite uncertain, needless to say, and are included only to give a rough idea of the times available for biological change.

Origin of the earth	4-5 billion years ago
Oldest records of bacteria and algae	3 billion
First filamentous algae	2 billion
Oldest nucleated cells	1 billion
Marine invertebrates	600 million
Fish, some land plants	400 million
Amphibians	300 million
Dinosaurs	200 million
Mammals	150 million
Primates	20 million

Suppose bacteria began 3 billion years ago and were able to divide about three times per day—say 1000 times per year. There would have been time for about 3,000 billion cell divisions, an opportunity for 3×10^{12} successive choices between two cells. This is an extension of the parlor game of 20 questions to 3,000 billion questions. Stated another way, if all the descendants of the original cell had survived, there would be $2^{3,000,000,000,000}$ today. With the capacity to discard all but one of these, there is an enormous opportunity to be selective.

MOLECULAR EVOLUTION

It has been abundantly clear throughout this book that the molecular understanding of genetics has enormously clarified our perceptions of almost every aspect of the subject. Population and evolutionary genetics are no exception.

An area in which our whole outlook has changed is the nature of genetic differences between species. The methodology of classical genetics is confined to breeding experiments, with some input from cytology. We know a great deal about the genetic control of the difference between a white mouse and a gray mouse, but we know nothing about the genetic basis of the differences between a mouse and a rat, to say nothing of those between a mouse and an elephant, or a man, or a dandelion. Knowledge of genetic differences has been confined to organisms where matings could be made. Molecular and cell genetics have changed this entirely. Cells from distantly related species that could never be crossed now can be fused in cell culture. More pertinent, for the purposes of this chapter, is the study of protein sequences, or even DNA in some cases, in different species. Not only can we determine how the genes differ in species that could never be crossed, we can even get information about genes in species that have long been extinct.

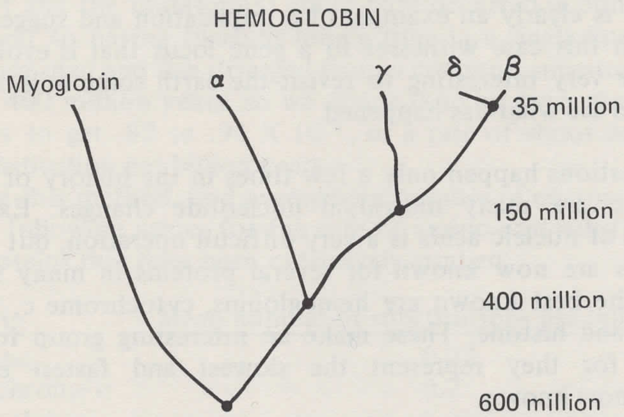
EVOLUTION BY GENE DUPLICATION

We discussed in Chapter 15 the possibility of gene duplication by an error in crossing-over. This could produce a repeat of two identical genes, one of which is free to evolve in new directions as the other carries out the original function. The best understood example of evolution by gene duplication is in the hemoglobins.

All the higher vertebrates have a hemoglobin that is a tetramer. The lamprey, however, has a hemoglobin that has only a single chain. Since the time when the ancestors of the lamprey split from those of the higher vertebrates, the hemoglobin-producing locus in the vertebrate line has duplicated, leading to the two loci producing alpha and beta chains. These two genes are now on independent chromosomes, so if they arose by unequal crossing-over they have since become separated, perhaps by translocation. Alternatively, there may have been a duplication of a whole chromosome through nondisjunction, or of a whole set of chromosomes through polyploidy. The latter seems more likely because of highly deleterious effects of chromosome unbalance.

There are now two other loci in man, closely linked to the beta locus: the delta and gamma hemoglobins. In their amino acid composition the beta and delta chains are very similar to each other, so they must have separated relatively recently. Both of these are less similar to gamma than to each other, but more similar to gamma than to alpha. To be more specific, human beta and delta differ by 10 amino acids, beta and gamma by 39, delta and gamma by 41, alpha and beta by 86, alpha and gamma by 87, and alpha and delta by 87.

The higher vertebrates all had a common ancestor 400 million years ago, so the split of alpha and beta chains must have been before that time. It cannot have been too much earlier, however, because the separation of these from the ancestors of the lamprey was only shortly before. Going still further back there was only one globin, which split by gene duplication into one gene responsible for oxygen storage, myoglobin, and one for oxygen transport, hemoglobin. Assuming that the number of differences between two proteins is roughly proportional to the time since their separation, we can put together a picture such as is shown.



We see in this example that five contemporary gene loci have evolved from a single ancestral gene. No other example is so well known, but this is a model of how we think that proteins with different functions evolve: by a duplication of the gene (perhaps as a part of a larger duplication of a group of genes, or even a whole chromosome), followed by divergence of functions as these follow their separate evolutionary courses.

Can we observe such changes occurring? The haptoglobins, serum proteins that have the function of binding loose hemoglobin in the blood, may provide an example. There are three haptoglobin types: 1F, 1S, and 2. In ordinary family studies these three behave as three alleles. By chemical analysis, 1F and 1S (F and S standing for fast and slow electrophoretic migration), differ only in one amino acid. However, the two 1's differ in a radical way from 2, haptoglobin 2 having almost twice the molecular weight of haptoglobin 1 and representing almost a complete duplication. It is two 1's joined together with a few amino acids missing at the junction. The missing DNA includes the terminating and starting codons, so that a single long molecule is produced by one continuous translation.

The two halves of the double gene are sufficiently similar that one would expect occasional out-of-register synapsis. Such an event would lead to single-length and triple-length genes. This suggests a hypothesis for maintaining the three kinds of haptoglobins in the population. Type 2 is for some reason slightly favored by natural selection. However, because of mis-synapsis, homozygous type 2 produces type 1's and triples. The hypothesis is that the rate of production of 1's just balances the selective advantage of 2 so that there is a stable polymorphism.

What about the predicted triples? They have in fact been found, but only rarely. Although duplicating the gene seems to produce a better product, triplicating it makes something worse. The triplet gene appears to be at an appreciable selective disadvantage although it causes no overt abnormality or disease.

Whether this hypothesis for the maintenance of the three types of haptoglobins in the population is correct, we don't know. But the gene is clearly an example of a duplication and suggests that we are in this case witnesses to a gene locus that is evolving. It would be very interesting to revisit the earth some million years hence and see what has happened.

AMINO ACID SUBSTITUTIONS IN EVOLUTION

Duplications happen only a few times in the history of a gene, but there are many individual nucleotide changes. Exact sequencing of nucleic acids is a very difficult operation, but protein sequences are now known for several proteins in many species. Among the best known are hemoglobins, cytochrome c, fibrinopeptide, and histone. These make an interesting group for comparison, for they represent the slowest and fastest evolving proteins now known.

Hemoglobin alpha has 141 amino acids. The upper right part of Table 25-1 gives the number of amino acids among these 141 that are different in human, horse, rabbit, mouse and fish (carp). As expected, the number of differences among the mammals is much less than the number between any of these and the fish. These figures will be more meaningful if we convert them into rates of change.

Table 25-1

	Man	Horse	Rabbit	Mouse	Carp
Man		18	25	17	58
Horse	.14		25	23	67
Rabbit	.20	.20		28	72
Mouse	.13	.18	.22		68
Carp	.66	.64	.71	.66	

First, we need to correct for failure to detect some intermediates. If one species has amino acid A and another has amino acid C, it may be that the ancestral amino acid A and that one line of descent since the split changed to C, or the ancestor could be C and one line changed to A; in either case there would be one substitution. However, it is also possible that the ancestor was B and changed to A in one line and C in the other. Usually, we would have no way of knowing that this had happened. Sometimes we can tell by the code; for example, it requires two nucleotide replacements to change methionine into proline. Most of the time, however, we rely on a purely statistical correction by assuming that the changes are independent. The lower left figures give the mean number of changes per codon, K, after such a correction.

Next, we want to convert these into the number of changes per codon *per year*. To do this we divide K by 2T, where T is the number of years since these two species diverged from a common ancestor. The reason for the 2 is that there are two lineages since the split and the mutation(s) could have occurred in either; this is equivalent to having twice as long a time in a single lineage. The time since man and fish diverged from a common ancestor is about 350 to 400 million years, so we divide 0.66 by twice this number of years to get .82 to $.94 \times 10^{-9}$, or a rate of about one amino acid substitution per billion years.

Using this method, and averaging a number of comparisons, we get the following values for the rate of amino acid substitution in four proteins that have been extensively studied.

Protein	Mean number of substitutions per billion years
Histone	0.02
Cytochrome c	0.2
Hemoglobin	1
Fibrinopeptide	4

How can we understand the large differences in rates for these four proteins? The most plausible explanation is that the proteins differ greatly in the extent to which particular amino acids at particular sites are necessary for proper function. If we look more carefully at the individual amino acid sites in hemoglobin, taking into account the three-dimensional structure brought about by folding the polypeptide chain, some patterns emerge. Those amino acids that are closely associated with the heme group have hardly changed at all. Presumably only these particular amino acids have the right properties to play their particular role and they cannot be changed without a serious impairment of the oxygen-carrying function. We know, however, that these codons are capable of mutation, because such mutants have been found in humans. But, they cause severe disease, and thus support the idea that these are sites where the molecule cannot be tampered with.

On the other hand, the amino acids that are on the outside of the molecule seem, for the most part, to be those that change most rapidly. This makes sense. These are less essential for oxygen-carrying and the molecule can still function if there are replacements at these sites. Presumably these replacements, although their effects are very minor, may cause slight changes that increase or decrease the effectiveness of the molecule—they are the fine tuning adjustments. If they improve the function, the causative genes are increased by natural selection. If they impair it, they are eliminated.

A molecule like cytochrome is more rigid in its demands on its component parts. More of the amino acids are irreplaceable, so a smaller fraction of sites permit substitution. The cytochromes offer one advantage for evolutionary study: since they evolve so slowly they can be used to study differences of very diverse groups, such as animals versus plants.

The extreme example of slow evolution is histone. The most studied histone fraction has 104 amino acids. Comparison of cattle and the pea plant showed only two changes. Animals and plants must have separated at least a billion years ago, so if only 2/104 of the amino acids changed, the average rate is less than 2×10^{-11}, or $.02 \times 10^{-9}$, as given above. Histone is closely associated with DNA in the nucleus. Its function must depend very strongly on its exact amino acid sequence.

At the other extreme is fibrinopeptide, which evolves about four times as fast as hemoglobin. Fibrinopeptide is a part of the blood clotting molecule, fibrinogen. When the clot is formed this part is broken off and not used. It is not surprising that an exact structure is not required, and the molecule is therefore free to change.

EVOLUTION BY RANDOM GENE FREQUENCY DRIFT

The fact that the same protein evolves about the same rate in widely diverse phylogenies has caused a great deal of speculation. One more example will help bring out the point. The separation of alpha and beta hemoglobin occurred some 400-500 million years ago. The ancestry of fish and man separated some 350-400 million years ago, so carp and man have been separated almost as long as the two types of hemoglobin. The corrected number of changes per codon, K, for three meaningful comparisons are:

Human alpha vs. carp alpha	K = .665
Human beta vs. human alpha	K = .776
Human beta vs. carp alpha	K = .807

The alpha chain has not evolved as fast as the beta in general, so we are not surprised that the first value is smaller than the other two. But the remarkable comparison is between the last two. The human beta and alpha, two proteins in the same organism—indeed in the same tetrameric molecule—have evolved differences from each other that are not significantly less than the differences between one of these and a molecule in a fish. Fish and humans have been separated from each other for a time almost as long as that since the original duplication of the original hemoglobin locus. The environments of humans and fish are about as different as we can imagine over a large part of their history. Fish have changed relatively little in superficial morphology during this time whereas the mammals and their ancestors have changed enormously. We must conclude that there is very little relationship between the rate of molecular evolution, at least in those molecules that have been studied, and external form and function.

This, and some other seeming paradoxes led M. Kimura in Japan and J. King and T. Jukes in the United States, quite independently from opposite shores of the Pacific, to make the daring suggestion that amino acid changes of the type we have been talking about are not the result of natural selection, but of mutation and random drift. King and Jukes called this "non-Darwinian evolution." The idea has been highly controversial. No one believes that all amino acid changes are random. But on the other hand it is hard to rule out the possibility that some are, and there is the possibility that this is a substantial fraction of all substitutions.

RAPID EVOLUTION

Ordinarily the evolutionary process is so slow that no significant change is observed in a human lifetime. However, there have been a few conspicuous exceptions, in each case the result of a radically changed environment.

One of these is **industrial melanism**. Some species of moths in England have gradually become darker in a period of a few decades. This period coincided with the increasing industrialization of the area and the greater production of smoke. These moths have become darker as the tree trunks were darkened with smoke and changing varieties of lichens. That there is actually strong selection for the protectively colored varieties is shown by observing birds feeding on them. On dark trees the birds caught mainly the light moths, and vice versa. Calculations show that in order to account for the change in gene frequencies very strong selection, of the order of 50%, must have taken place.

Quite similar examples are provided by the rapid acquisition of resistance to DDT and other organic insecticides. Here too, there was a strong selective pressure in favor of the resistant mutants, and populations became highly resistant in a few years. Laboratory reconstruction experiments have demonstrated that natural selection is a sufficient explanation. That no Lamarckian explanation is needed is shown by indirect selection; that is, selection of flies whose sibs were resistant. In this way a population of resistant flies was developed whose ancestors had never been exposed to the insecticide. Drug resistance in bacteria follows the same principles.

These examples demonstrate that evolution can under some circumstances proceed very rapidly, the necessary circumstance being a radical change in the environment so that different genetic types are favored by selection.

Among evolutionary rates that have not been influenced by human intervention, one of the most rapid has been that of the human skull. It is much more rapid than other changes among the mammals, which in turn have been more rapid than other vertebrates. (Vertebrate evolution, in turn, has occurred more rapidly than most in invertebrates and plants.) Presumably, this has been the result of natural selection for a larger brain. The development of higher intelligence must have been of consistent selective value, for this increase in size has been gotten at a high cost—the increasing difficulty and danger of the birth process.

EVOLUTIONARY ADVANTAGES OF MENDELIAN HEREDITY

If evolution depends on polygenic traits, as it surely must to a very large extent, there is a clear advantage of recombination. If there is a change in the environment so that, for example, an increase in size is desirable, an asexual population can respond only by increasing the frequency of the largest type in the existing population. But with recombination, the existing genes can be reassorted to produce types that are much larger than any type now present. This has been demonstrated time and again with selection experiments where types far removed from even the most extreme members of the original population have been produced by selection, such as mice that are almost as large as small rats.

Recombination also offers an advantage for incorporating new favorable mutants into the population. In an asexual population two beneficial mutants can be incorporated into the population only if the second occurs in a descendant of the individual in which the first occurred. Thus the mutants can be incorporated only one at a time, in series.

In a sexual population, on the other hand, the various mutants present in the population can get into the same individual by recombination. Only if the mutation rate were so low or the population so small that each mutant became established before the next favorable mutant occurred would the sexual system have no advantage over the asexual.

This is true of genes that are beneficial both singly and in combination. (We are ignoring harmful mutants which are eliminated from the population and make no evolutionary contribution.) The situation is quite different if there is strong interaction between the mutant genes. When two or more mutants are individually harmful, but beneficial in combination, sexual reproduction may be actually disadvantageous. For, if the individual with the two mutants arises, the gene combination is broken up by recombination leading to the harmful types. Suppose in a haploid organism that $+ +$ is the standard form, $A +$ is harmful, $+ B$ is harmful, but that the double mutant $A B$ is highly advantageous, better than the standard form. With a sexual species there is no way to go from $+ +$ to $A B$ since the intermediate recombinant forms are disadvantageous. On the other hand, in an asexual species there is no recombination and $A B$ type, once it arises, can be increased by selection and ultimately becomes the prevailing type.

This analysis would suggest that Mendelian inheritance is advantageous if genes work more or less cumulatively, but not if specific interactions are a major factor in evolution. There are obvious immediate advantages to asexual reproduction. If the heterozygote is advantageous, there is not the problem of inferior homozygotes being produced by Mendelian segregation. All this fits with the conventional belief that sexuality developed very early in evolution so that it is found in all major groups; but that numerous independent retrogressions to vegetative reproduction continue to occur, conferring an immediate advantage but a long term evolutionary disadvantage. Most of these eventually become extinct as the environment changes.

EVOLUTION OF ALTRUISTIC AND COOPERATIVE BEHAVIOR

It is easy to see how evolution produces stronger, healthier, and more competitive individuals. On the other hand, it is not so easy to see how cooperative social behavior evolves. In particular, it is hard to see how evolution of self-sacrificing or altruistic behavior can arise by a system of competitive natural selection.

A great many seemingly altruistic patterns of behavior may not really be this. Some apparent altruism may actually benefit the individual in some way, perhaps not an obvious one. Also a great deal of altruism may really be cooperative and benefit both parties; I make a minor sacrifice to help you, you do the same for me, we both gain more than we lose.

Truly self-sacrificing behavior is most easily explained by "kin selection." An extreme example of self-sacrifice is the worker bee, who risks being killed every time she stings. This can be understood by noting that the worker bee is sterile, and the only way that she can hope for posterity is by keeping the queen alive. This perpetuates some of the worker's genes, for the worker and queen share more than 50% of their genes.

In a less extreme way, we can understand that selection would lead to behavior patterns whereby a mother protects her child, even at considerable sacrifice to herself. Since she and the child share 50% of their genes she might be expected rationally (from the standpoint of natural selection) to run a risk up to 50% if this assures the survival of the child. Also, the child has its whole reproductive career ahead, whereas the mother may be near the end of hers. On the other hand it is not worth much of a sacrifice if the offspring is likely to die of other causes anyhow. Therefore we would expect much more protective behavior in mammals than in insects, where most of the progeny regularly die at an early age.

The structure of human populations in the past was one of division into many small groups, rather isolated from each other. Individuals within a group would all tend to be somewhat related to each other, and to differ from other groups. The same principle would then lead natural selection to favor cooperative behavior in the group, especially as regards its relations to other groups. The important role that this form of natural selection may have played in the development of human social systems has been emphasized especially strongly by H. J. Muller.

ORIGIN OF SPECIES

Sometimes a population of individuals will become separated into two or more groups, perhaps by some geographical barrier such as a mountain range or a body of water. If the two groups are in different environments different gene combinations will be selected in each group so that they will gradually diverge. The same thing will happen by chance, though perhaps more slowly, even if the environments are identical. After a sufficient period of time the isolated populations will have become so different that they are no longer inter-fertile. Now they have become separate species, and there will be no gene flow between them even if they later occupy the same area.

The process of origin of species is thus one of slow divergence of isolated populations over many generations. If they were not isolated, the accumulated differences would be broken down by hybridization; but as long as they are isolated, by geography or later by cross sterility, they can follow separate evolutionary courses. The main and perhaps the only exception to this slow process of species formation is the production of a new species in a single generation by polyploidy. Through chromosome doubling, plants fertile among themselves but not with their diploid parents are formed; hence they are immediately isolated.

That situations which offer chances for geographical isolation favor speciation is evidenced by the rapid speciation that takes place in some island groups. That geographical isolation can cause speciation and that much speciation occurred this way is generally agreed. But whether *all* speciation (except by polyploidy) takes place through geographical separation is now a topic of lively debate among students of evolution.

26

THE ORIGIN OF LIFE

WHAT IS LIFE?

The borderline between the living and the nonliving is not at all clear. A streptomycin-treated F^+ bacterial cell cannot divide (since this requires protein synthesis), but it can serve as a chromosome donor by mating. On the other hand, if capacity to reproduce is the criterion, *one* rabbit is not living. Alternatively, a properly punched IBM card is self-reproducing in some sense. Is a crystallized virus living or not? Modern chemistry of DNA and proteins is bridging the gap between test tube chemistry and living cells.

By now, you have probably concluded (with me) that this is a sterile question. So, as the topic for this chapter, I will try to explain how cells might arise from inorganic matter. The exact point where it might have been called "living" is inconsequential.

The basic problem is the origin of a genetic system. As long ago as 1921 H. J. Muller argued that life depends not only on self-reproduction, but on the ability to multiply variants; that is to say, on self-replication and mutation. Then natural selection is possible.

AUTOTROPH VERSUS HETEROTROPH THEORY

An **autotroph** is an organism that is nutritionally nonexacting: it will grow on very simple chemicals. A **heterotroph** requires complex organic compounds.

Traditionally, it has been thought that the earliest organisms were autotrophs; for organic compounds come from living organisms. No living organisms, no organic compounds. But living forms also eat or otherwise destroy organic compounds. J. B. S. Haldane was the first to point out that, in the absence of organisms to destroy them, organic compounds, if they were formed, would accumulate until primitive waters "reached the consistency of hot dilute soup." Hence, the earliest precursors of life may have had an abundance of organic material. This idea, developed by Haldane and Oparin, is now generally accepted.

It is interesting to note that Charles Darwin was pretty close to the mark. In a letter he wrote: "It is often said that all the conditions for the first production of a living organism are now present, which could ever have been present. But if (and oh! what a big if!) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc., present, that a proteine compound was formed ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed."

THE PRE-BIOTIC ATMOSPHERE

The present atmosphere is oxidizing: it includes nitrogen, free oxygen, carbon dioxide, and water. There is very good evidence that the early atmosphere did not contain any free oxygen: based on observations of other planets, it is thought that the prevailing compounds would be methane (CH_4), ammonia (NH_3), water, and free hydrogen—a strongly reducing environment. Some have argued that the primitive atmosphere was more nearly neutral, containing CO_2 and free nitrogen along with water and ammonia; this conclusion being based on studies of volcanic eruptions. There is complete agreement, however, that there was no free oxygen.

The absence of free oxygen has a significant consequence. At present most ultraviolet irradiation is filtered out by oxygen, especially in the form of ozone. In the absence of atmospheric oxygen there was much more energy available for chemical reactions activated by UV then than now. There was also energy from electricity (presumably there were thunderstorms and lightning then, as now), and from radioactivity.

Our present atmosphere with its large amount of free oxygen is the result of plants' life processes. Rocks that were formed prior to the "invention" of photosynthesis are in a relatively reduced state, while later ones are more oxidized.

THE MILLER EXPERIMENT AND ITS SUCCESSORS

The key experiment which proved that organic compounds could have been synthesized from inorganic materials was performed in 1953 by Stanley Miller, then a graduate student at the University of Chicago. It was as simple in concept and technique as it was revolutionary in its implications. A mixture of methane, hydrogen, ammonia and water vapor was circulated through a vessel in which electrodes were producing a spark discharge. The products accumulated in a small heated flask from which volatile compounds were continuously distilled and passed by the spark. It was an attempt to simulate conditions that might have prevailed in the primitive earth.

At the end of a week (not a billion years, as some would have predicted), about 5% of the carbon had been converted into organic compounds. The most striking feature is that virtually all the compounds produced were biologically important, especially the amino acids. Hydrogen cyanide and formaldehyde were also formed. The significance of the former is that this together with ammonia readily yields adenine. Subsequent experiments by various workers have shown that virtually all the amino acids and nucleic acid components, and precursors of such important compounds as ATP, are produced in such systems. The source of energy, whether heat, ultraviolet radiation, ionizing radiation, or electric discharge, doesn't seem to make much difference.

(The reaction that produces adenine, incidentally, seems to go best at refrigerator temperatures. Perhaps part of the process requires not Haldane's hot dilute soup, but cold concentrated soup.)

It is significant that this type of experiment does not yield amino acids except under strongly reducing conditions. It is also significant that the compounds yielded in experiments with reducing conditions are almost all commonplace, but essential biological molecules. It argues strongly that the present chemical composition of living organisms was largely determined by the kinds of organic compounds that were first present on the earth. What we are made of is what was there.

POLYMERIZATION

There are a number of plausible mechanisms that would favor the aggregations of small molecules into longer polymers. One simple possibility was suggested by Carl Sagan: ultraviolet radiation causes polymerization, but it also depolymerizes. If a large molecule were formed near the surface of a pool, it would tend to sink and be protected from the depolymerizing effects of the ultraviolet rays, since water is quite opaque to UV. Many other plausible mechanisms exist, so that one could expect that if monomers are produced there would be, somewhere on the earth, conditions for polymerization. Actual experiments have yielded nucleic acid polymers of a half-dozen nucleotides—not very long by present DNA standards, but perhaps long enough that some kind of primitive self-replication could begin.

The earliest molecules would not need to have a high probability of replication. It need be only high enough to offset thermodynamic decay. But if molecules can replicate at all, even with very low probability, then those that are the least bit better at this process than others will have a selective advantage. And there was a lot of time.

CONCENTRATION

Those molecules (perhaps nucleic acids) that could make copies of themselves most efficiently probably did so by somehow cooperating with other molecules, perhaps amino acids. If this were true, then those molecules that could concentrate a halo of other molecules that would expedite the self-copying process would be at an advantage. Such a process could gradually lead to the highly efficient means that cells have for concentrating the necessary materials. I don't mean, of course, to suggest that anyone knows how this actually happened; but only that it doesn't strain our credulity that some such thing might have happened.

Various concentrating mechanisms exist: evaporation, absorption on the rocks and clays, and freezing are some possibilities.

ORIGIN OF BIO-SYNTHETIC PATHWAYS

Present day organisms carry out long sequential steps of biochemical transformations. These must have evolved by small steps, each with a selective advantage. It is also striking that, although the earliest forms of life must surely have been heterotrophs, many of the present-day lower plants are autotrophic. How did these synthetic processes evolve? A specific scheme was suggested by Norman Horowitz.

Suppose there are a series of chemical compounds existing in the environment. The chemical properties of these compounds are such that A is potentially capable of being converted into B, B into C, C into D, D into E, and E into F (for food), which is needed by the organism. These compounds are all assumed to be present in the environment, perhaps in some kind of purely chemically-determined equilibrium. Now when the supply of F is exhausted, there will be a selective advantage to any mutation that enables the organism to convert E into F, and this type will soon supplant the others. Then the supply of E will be exhausted, and a mutant that converts D into E (while still retaining the ability to convert E into F) will be favored. And so on, so the whole chemical conversion process $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow F$ is acquired in the backwards direction.

LIFE ON OTHER PLANETS

Astronomers tell us that there are millions of other planets which have conditions similar enough to those of the earth for life to be possible. The discussion up to now has intended to convey the idea that life is not an improbable concatenation of events, but an orderly sequence of steps each with a reasonably high probability. If life is really a highly probable thing, then we may expect it to be widely dispersed in the universe. Mars still looks like the best possibility in our own solar system.

There is now very good evidence that meteorites contain organic compounds. The suspicion that these might be the result of contamination with earthly microorganisms is largely dispelled by finding that the amino acids exist in racemic mixtures rather than in only one stereoisomer such as is characteristic of life on earth. Furthermore, some of the amino acid-like compounds found in meteors are not prevalent in living organisms; also, there is more C^{13} in meteorites than in products of living cells.

Can life be transported from one planet to another, in the form of spores perhaps? Or did life originate independently on several planets? If so, are DNA and RNA the basic chemicals of life elsewhere as they are here? The question of whether self-reproducing systems not involving nucleic acids can exist is most intriguing. Surely one of the most basic questions in biology is this: is the primacy of DNA a chemical necessity or just a historical-biological accident?

THE CHI-SQUARE METHOD

Appendix I

SIGNIFICANCE TESTS

A question that constantly arises in the study of genetic ratios is: how far from the theoretical expectation must the observed results be before they can be regarded as indicating a real difference and not just an accident of sampling? There is, of course, no exact answer to this question. You can never be absolutely sure that the results were not due to chance no matter how large the difference.

But in practice one has to make decisions without being absolutely sure. The experimenter has to decide, for example, whether the data can be considered to be in sufficiently good agreement with a 3:1 expectation or not, and then act on the basis of this decision. There are two kinds of mistakes a researcher can make: (1) decide that the true ratio in the character under study is not 3:1 when it really is, or (2) decide that the true ratio is 3:1 when it really isn't. (These are examples of what a statistician calls type I and type II errors.) The purpose of statistical procedures is to determine the likelihood of errors, and, if possible, to minimize them.

Suppose an experimenter tests 100 backcrossed corn plants for a 1:1 ratio. The result is 65 green and 35 yellow plants. Should he conclude that his 1:1 expectation is wrong and that he must be dealing with some more complex situation than simple monohybrid inheritance, or did the results just happen by chance to be this far off? One procedure that he can follow is to ask the question: what is the probability of getting results this bad or worse if the expectation *is* correct? The probability of getting results differing from the expected 50:50 by 15 or more can be computed, and after much tedious work he finds it to be about 1 in 300. With a probability this small the experimenter would probably conclude that the 1:1 expectation was incorrect, but this could still be the 1 chance in 300 that might happen. On the other hand, if the probability had turned out to be 1 in 10, he would have concluded that this might very well be an accident of sampling, and that there was no reason on the basis of these data for rejecting the 1:1 hypothesis. (This, of course, does not verify the 1:1 hypothesis; it just fails to provide evidence against it.)

This is the kind of analysis that an experimenter makes when performing a **significance test** of a hypothesis. How small the probability must be before he rejects his hypothesis is arbitrary and depends on the seriousness of the consequences of an error, but it is conventional to regard 1/20 or 1/100 as borderline probabilities.

THE CHI-SQUARE METHOD

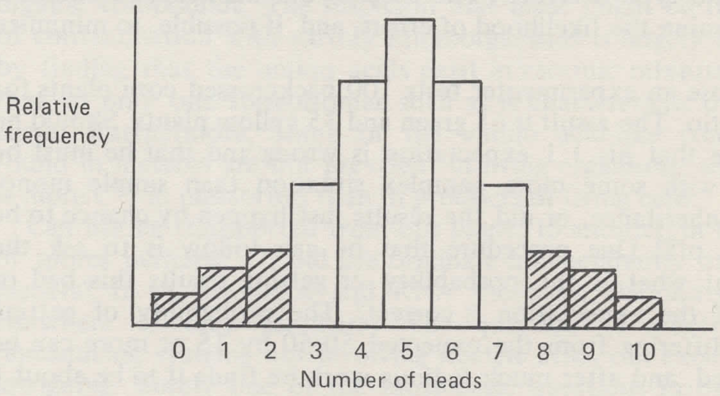
As the number of individuals in a sample becomes large, the arithmetic of computing probabilities by the methods of the last section becomes increasingly tiresome. In this section a method of getting approximate answers will be developed which requires much less computation and which gives answers accurate enough for most purposes. The method was given by Karl Pearson in 1900 and is known as the chi-square method.

An illustration that will give some idea of the nature of the approximation involved can be given. As an example consider the probabilities of throwing 0, 1, 2, . . . 10 heads in a toss of 10 coins. The following results are obtained by the method of Chapter 5.

Table I - 1

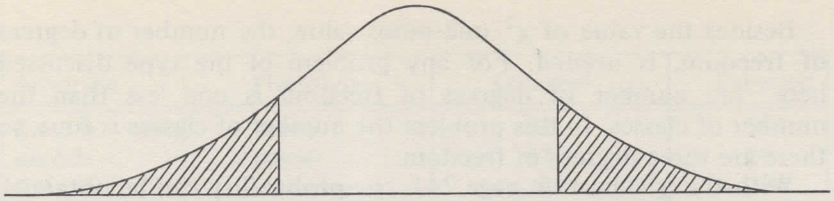
Number heads	0	1	2	3	4	5	6	7	8	9	10
Probability of occurrence	$\frac{1}{1024}$	$\frac{10}{1024}$	$\frac{45}{1024}$	$\frac{120}{1024}$	$\frac{210}{1024}$	$\frac{252}{1024}$	$\frac{210}{1024}$	$\frac{120}{1024}$	$\frac{45}{1024}$	$\frac{10}{1024}$	$\frac{1}{1024}$

If this were diagrammed with the number of heads as abscissa (horizontal axis) and the frequency of occurrence as ordinate (vertical axis), we should obtain the figure shown below. Such a graph is called a histogram.



The cross-hatched area divided by the whole area of the graph would represent the probability of differing from the expected number by 3 or more.

If the number of coins tossed were 100 instead of 10, the size of the individual stair-steps of the histogram would be much smaller, and as the number of coins tossed is increased the graph becomes more and more similar to a smooth, bell-shaped curve. The desired probability would be given by the shaded area divided by the total area.



The chi-square approximation involves using such a smooth curve as a substitute for the exact histogram. If the number of individuals in the sample is large, the individual stair-steps are small and the approximation is very good. On the other hand, if the number is small, the stair-steps are large and the approximation is not so good. Thus the chi-square method is better when the number of individuals in the sample is large. This is fortunate, for when the number is large, the exact computation becomes prohibitively difficult.

The area under a portion of the smooth curve is found by application of the calculus. Fortunately this has already been done and the results placed in table or graph form, so that it is only necessary to read the table or graph to obtain a desired probability. The chi-square procedure is simply a device by which approximate probabilities can be found by simple procedures when the exact method would be too time consuming.

The value of χ^2 is given by

$$\chi^2 = \frac{\Sigma(\text{observed - expected})^2}{\text{expected}} \qquad \text{(I-1)}$$

where the symbol Σ stands for the sum of all the classes. This value is then used with a chart or table to obtain the probability. This can best be explained with an example.

EXAMPLE

In some matings involving different color phases in the mink, Shackelford (*Am. Naturalist* 83:49) found 53 dark, 17 platinum, 24 aleutian, and 8 sapphire offspring from a series of matings. He had theoretical reasons to suspect that there were two genes involved and that a 9:3:3:1 ratio was to be expected. Were the results in agreement with this hypothesis?

Table I-2

Class	Expected proportion	Observed number	Expected number
Dark	9/16	53	57.4
Platinum	3/16	17	19.1
Aleutian	3/16	24	19.1
Sapphire	1/16	8	6.4
Total		102	102.0

$$\chi^2 = \frac{(53-57.4)^2}{57.4} + \frac{(17-19.1)^2}{19.1} + \frac{(24-19.1)^2}{19.1} + \frac{(8-6.4)^2}{6.4} = 2.23$$

Besides the value of χ^2 one more value, the number of **degrees of freedom**, is needed. For any problem of the type discussed here, the number of degrees of freedom is one less than the number of classes. In this problem the number of classes is four, so there are three degrees of freedom.

With the χ^2 chart on page 241, the probability can be obtained. (Look along the lower horizontal axis until the value of χ^2 (in this case 2.23) is found. Then go up from this point to the line corresponding to the number of degrees of freedom (in this case 3). Directly to the left is the probability, about .55.

INTERPRETATION OF THE PROBABILITY

The probability given by the chart is the probability of obtaining by chance results that differ from the expected results by as much as or more than those results actually observed, if the hypothesis is correct.

If the probability is high, there is good agreement between expectation and observation. If it is small, the agreement is poor. If the probability is very small, we decide that the data do not support the hypothesis and reject it. The probability level at which the hypothesis leading to the expected results is to be rejected must, of course, be arbitrary. However, the following conventions are usually followed:

1. If the probability is greater than .05, the data are generally regarded as in agreement with the expectation. The differences between observation and expectation might well be due to chance alone. (Note that this does not prove that the hypothesis is correct; it merely fails to offer significant evidence that it is incorrect.)
2. If the probability is less than .05 but greater than .01, the observations are said to differ significantly from the expectation, or the difference is said to be *significant.*
3. If the probability is less than .01, the difference is said to be *highly significant*. In this case the data are not in agreement with the expectation, and the hypothesis is considered as disproved. But we must remember that if we continue to reject hypotheses at the 1% level, we must expect to reject a correct hypothesis 1% of the time.

In the example above, the probability of .55 means that there was good agreement between observation and hypothesis. There is no reason, on the basis of this data, for rejecting a 9:3:3:1 hypothesis.

A SECOND EXAMPLE

Another mink cross involving the genes ebony (E) and pastel (b) was made between $E/e\ B/b$ and $e/e\ b/b$. See data on opposite page.

In this case the probability of getting by chance results this far away from the expected results is less than one in 10,000. This is a highly significant difference, and the hypothesis leading to a 1:1:1:1 ratio was rejected. (In this case the explanation is that the E and B genes are carried on the same chromosome, and hence are not inherited independently.)

Table I-3

Genotype	Phenotype	Observed	Expected
E/e B/b	ebony	22	20
e/e b/b	pastel	37	20
E/e b/b	palomino	7	20
e/e B/b	dark	14	20
Total		80	80

$X^2 = 24.9$, Degrees of freedom = 3, Probability $< .0001$

REDUCED DEGREES OF FREEDOM WHEN SOME QUANTITIES ARE ESTIMATED FROM THE DATA

The examples so far used expected proportions based on some theory, Mendelism in these cases. It is also possible that the expected numbers depend also on some quantity or quantities estimated from the data themselves. A good example is the test of Hardy-Weinberg ratios for the MN blood groups in Chapter 20. The expected numbers are based on the random mating assumption, but they involve determining the gene frequency p from the data.

The rule is simple: Subtract one degree of freedom for each *independent* quantity estimated from the data.

In the example, we used the data to estimate p, the frequency of *M* genes from the data. We also estimated q, the frequency of *N* genes, but this is not an independent estimate, since $p + q = 1$.

PROBLEMS

1. Test the following observed results for agreement with a 1:1 expectation.
   a. 60 yellow, 40 green (difference of 10 or 10%).
   b. 510 yellow, 490 green (difference of 10 or 1%).
   c. 600 yellow, 400 green (difference of 100 or 10%).
   What generalization can you make about the effect of increasing sample size?
2. Test the following for agreement with a 9:3:3:1 ratio: 66, 22, 30, and 10.
3. The mink data above can be further subdivided. Compute chi-squared for the ratio of *E/e* to *e/e* (ignoring the B factor). Then compute chi-squared for the ratio of *B/b* to *b/b*. Finally, you can test for the independence of the factors by comparing the ratio of *E/e B/b* + *e/e b/b* to *E/e b/b* + *e/e B/b*. Where does the serious discrepancy lie? (You can check your computation by seeing if these three chi-squares, each for one degree of freedom, add up to the previously computed total chi-square for three degrees of freedom. This illustrates the additivity property of chi-square.)
4. Suppose that in an experiment a proportion p of yellow and 1 − p green are expected. Out of a total number N examined, a proportion a are observed to be yellow and 1 − a to be green. Show that the square root of χ^2 is

$$\chi = \frac{a - p}{s}$$

where

$$s^2 = \frac{p(1 - p)}{N}$$

(The quantity s is called the standard deviation..Notice that in the chart the probabilities for χ^2 with one degree of freedom correspond to those for t with infinite degrees of freedom; in this case, $\chi = t$. So for χ^2 problems with one degree of freedom you can use this formula instead of the standard formula, and the t chart instead of the χ^2.)

5. What is the value of χ^2 for one degree of freedom corresponding to a probability of .05? .01? What are the values for 10 degrees of freedom? What are the values of t for an infinite number of degrees of freedom? If you had a χ^2 chart, but no t chart, how could you determine these t values?
6. What are the χ^2 values corresponding to a probability of 0.5 for 10, 20, and 30 degrees of freedom? Do these suggest any generalization?
7. If *A* is the observed number of recessive homozygotes in an intercross when 1/4 is the expected proportion and N is the total number, derive a formula for χ^2. Derive a similar formula in terms of *a*, the observed proportion of recessives.
8. If the observed numbers in two classes are *a* and *b* and these classes are expected in the ratio *r:s*, show that

$$\chi^2 = \frac{(as - br)^2}{rs\,(a + b)}$$

9. Verify the χ^2 values in the first two tables on pp. 164-65 in Chapter 20. Notice the reduction in degrees of freedom because of quantities estimated from the data.

Notes:

1. These examples emphasize the rule that the standard formula for computing χ^2 apply to numbers, not to proportions; so be sure to use the numbers, observed and expected, in doing the calculations, not the proportions or percentages.
2. You might find it helpful to darken or otherwise emphasize the lines on the χ^2 and t charts corresponding to .01 and .05. The values falling above the .01 line (i.e., corresponding to a probability of less than .01) are "highly significant," those between .01 and .05 are "significant," and those below the .05 line are "not significant."
3. The χ^2 approximation is poor when the numbers are too small. The usual rule of thumb is not to use the χ^2 test when any expected number is less than 5.

Appendix II

CONFIDENCE LIMITS FOR BINOMIAL DISTRIBUTION

STATISTICAL ESTIMATION

In some genetics problems there is no theoretical expectation, but an unknown quantity is to be estimated from the data. This is true, for example, in linkage data. The ratios obtained from the data are used to estimate the map distance. When the true recombination value is to be estimated, it is desirable to have some idea of how accurate the estimate is. Insofar as the errors are random, their magnitude can be assessed by probability analysis and an estimate of the reliability of the estimate can be obtained.

FIDUCIAL OR CONFIDENCE LIMITS

We can use the χ^2 method of the previous chapter, but the other way around. Let p be the (unknown) true proportion that is to be estimated and o the observed proportion based on N observations. Then, from the usual formula for χ^2 for one degree of freedom,

$$\chi_1^2 = \frac{(\mathrm{No} - \mathrm{Np})^2}{\mathrm{Np}} + \frac{[\mathrm{N}(1 - \mathrm{o}) - \mathrm{N}(1 - \mathrm{p})]^2}{\mathrm{N}(1 - \mathrm{p})}$$

$$= \frac{\mathrm{N}(\mathrm{o} - \mathrm{p})^2}{\mathrm{p}(1 - \mathrm{p})} \tag{II-1}$$

For one degree of freedom, the value of χ^2 corresponding to a probability of .05 is 3.841 (verify that the chart gives approximately this value). Thus the probability that $\mathrm{N}(\mathrm{o} - \mathrm{p})^2/\mathrm{p}(1 - \mathrm{p})$ is equal to or greater than 3.841 is .05; or, conversely, the probability that this quantity is less than 3.841 is .95. Therefore we solve the quadratic equation

$$\mathrm{N}(\mathrm{o} - \mathrm{p})^2 = 3.841\ \mathrm{p}(1 - \mathrm{p})$$

and the two solutions for p are called the upper and lower 95% confidence limits. We can say that the probability is .95 that the true value lies between these limits, or .05 that it lies outside them.

(Whether the previous sentence is strictly correct and whether this is the proper use of the word *probability* has been the topic of many heated polemics, as has the exact distinction between fiducial and confidence limits. You may enjoy reading R. A. Fisher's *Statistical Methods and Scientific Inference* in this context.)

The value 3.841 is appropriate for 95% limits. Here is a short table of values for other confidence values.

Confidence limits	χ^2
.50	.455
.80	1.642
.90	2.706
.95	3.841
.98	5.412
.99	6.635

EXAMPLE

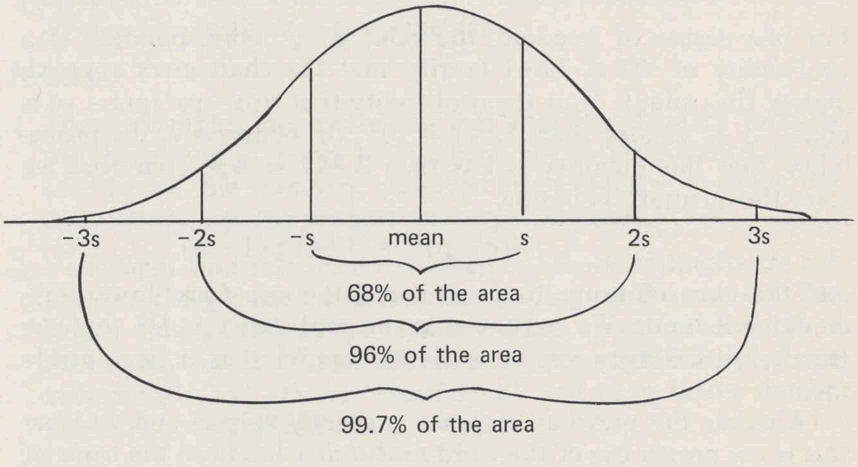
In the mink data in the previous chapter there were 21 recombinants among 80 progeny or 26.3%. Hence $o = .263$ and $N = 80$. So we solve the equation $80\,(.263 - p)^2 = 3.841\,p(1 - p)$. This quadratic equation has two solutions, .369 and .175, so these are the upper and lower 95% confidence limits.

THE STANDARD DEVIATION

If the data are binomial in nature, the standard deviation (s) expressed as a proportion is given by the formula

$$s = \sqrt{\frac{p\,q}{N}} \tag{II-2}$$

where p and q have their usual meaning and N is the total number counted. The standard deviation has the property that in a large number of samples approximately 68% of them would lie within one standard deviation of the mean, 96% would lie within two standard deviations, or 99.7% within three standard deviations as shown on the drawing below.



These percentages of area are only crude approximations if the sample size is small, but become increasingly accurate as the size of the sample increases.

APPROXIMATE CONFIDENCE LIMITS

The property of the standard deviation just described makes possible the setting of approximate confidence limits that, although not as accurate as those obtained by solving the χ^2 quadratic equation, are usually good enough. We make use of the fact that t for an infinite number of degrees of freedom is the square root of χ^2 for one degree of freedom (see problem 4 in the previous chapter).

We define the confidence limits as follows:

$$\overline{L} \text{ (the upper limit)} = p + ts$$
$$\underline{L} \text{ (the lower limit)} = p - ts \qquad \text{(II - 3)}$$

where t is chosen from the following table to give the required confidence limits.

Confidence Desired	t
50%	.67
90%	1.64
95%	1.96
99%	2.58

EXAMPLE

In the mink data in the last chapter there were 21 crossovers out of a total of 80 or 26.3%. Hence $p = .263$, $q = .737$, and $N = 80$. The standard deviation is given by $s = \sqrt{(.263)(.737)/80} = .053$.

For 95% confidence limits, $t = 1.96$. Therefore

$$\overline{L} = .263 + (1.96)(.053) = .366$$
$$\underline{L} = .263 - (1.96)(.053) = .160$$

and the 95% confidence interval is from 16.0% to 36.6% crossing-over.

When one makes the statement that the mean lies between $\underline{L}$ and $\overline{L}$ he has "95% confidence" (or 99% or some other value depending on the t value chosen) that this region includes the true value. That is, if a researcher continues to make such statements, then he or she can expect to be correct 95% of the time and wrong 5%.

Note that these values agree quite well with those obtained by the more troublesome method of solving the quadratic equation.

PROBLEMS

1. Write the solution for the quadratic equation $N(o - p)^2 = Cp(1 - p)$, where N, o, and C are constants.
2. An experimenter observes 10% recombinants among 100 back-cross plants. Use the equation from problem 1 to compute the 95% confidence limits.
3. What are the approximate 95% confidence limits.

Appendix III

SIGNIFICANCE TESTS WITH MEASUREMENT DATA

THE MEASUREMENT OF VARIATION

In examining the variability of a population one could measure the range of variation; that is the difference between the smallest and the largest. This is often a very useful procedure because of its simplicity. But it is easy to see that the value observed depends only on the magnitude of the two most extreme members of the sample studied. Since more extreme measurements are more likely to be found in a large sample than in a small one, the range is dependent not only on the inherent variability of the population but also on the number of individuals measured. Another undesirable feature is that, since its value depends on only two measurements, it is more likely to be subject to chance fluctuations from sample to sample and is hence less reliable in a statistical sense.

Any population whose distribution of values can be plotted to form a roughly symmetrical, bell-shaped curve lends itself to relatively simple statistical procedures. Such a population is normally distributed and a great number of simple and effective procedures have been worked out for handling quantitatively such measurements. The following statistics are among the most useful.

THE MEAN

The mean, $\bar{x}$ is the sum of all the measurements divided by the number of measurements made. Symbolically it may be written,

$$\bar{x} = \frac{\Sigma x}{N} \qquad \text{(III - 1)}$$

where
- x stands for each individual measurement
- N is the number of measurements made
- Σ means the sum

THE VARIANCE

The variance is a measure of variability of the population. It is estimated from the sample measurements by this formula

$$\text{Variance} = V = \frac{\Sigma[\,(x - \bar{x})^2\,]}{N - 1} = \frac{\Sigma(x^2) - N(\bar{x}^2)}{N - 1} \qquad \text{(III-2)}$$

One of the most useful properties of the variance is that it can be broken up into individual components. The process of dividing the variance into components and associating these components with experimental effects is known as analysis of variance and is widely applied in scientific research. See Chapter 24.

THE STANDARD DEVIATION

The standard deviation, s, is the square root of the variance. It has the same meaning as in connection with enumeration data discussed previously.

The standard deviation of the mean is the standard deviation divided by the square root of the number in the sample. Thus its formula is

$$s_{\overline{x}} = \sqrt{\frac{1}{N(N-1)} \quad [(x^2) - N(\overline{x}^2)]}$$

CONFIDENCE LIMITS ABOUT THE MEAN

$$\overline{L} = \overline{x} + t\, s_{\overline{x}}$$

$$\overline{L} = \overline{x} + t\, s_{\overline{x}}$$

The value of t is obtained from the t chart by choosing the appropriate probability with degrees of freedom one less than the number of measurements in the sample. For a confidence interval of 95% choose the probability .05; for 99% limits, use .01, etc.

TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO MEANS

The procedure and interpretation is much the same as in a χ^2 test. The first thing to do is to compute a quantity, t, defined as

$$t = \frac{\overline{x}_a - \overline{x}_b}{\sqrt{\dfrac{(N_a + N_b)\,(\Sigma x_a^2 + \Sigma x_b^2 - N_a\overline{x}_a^2 - N_b\overline{x}_b^2)}{N_a N_b (N_a + N_b - 2)}}}$$

$$= \frac{\overline{x}_a - \overline{x}_b}{\sqrt{\dfrac{(N_a + N_b)\,[(N_a - 1)V_a + (N_b - 1)V_b\,]}{N_a N_b (N_a + N_b - 2)}}}$$

$$\text{Degrees of freedom} = N_a + N_b - 2$$

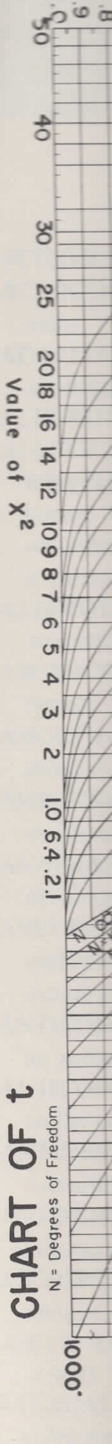
The subscripts refer to samples a and b.

The probability is read from the t chart just as it is from a χ^2 chart. Just turn the chart upside down and follow the same procedure. If the probability is less than .05 the means are regarded as significantly different. If the probability is less than .01, the difference is highly significant.

PROBLEMS

1. A set of measurements are 3, 10, 11, 8, 17, 9, 15, 13, 5, 7, 12, 10. Compute (a) the mean, (b) the variance, (c) the standard deviation, (d) the standard deviation of the mean, (e) the 95% confidence limits about the mean.
2. Is the difference between the following two strains of corn significant?

Strain	Number Measured	Mean Height (cm.)	Variance
A	38	197	910
B	49	218	712



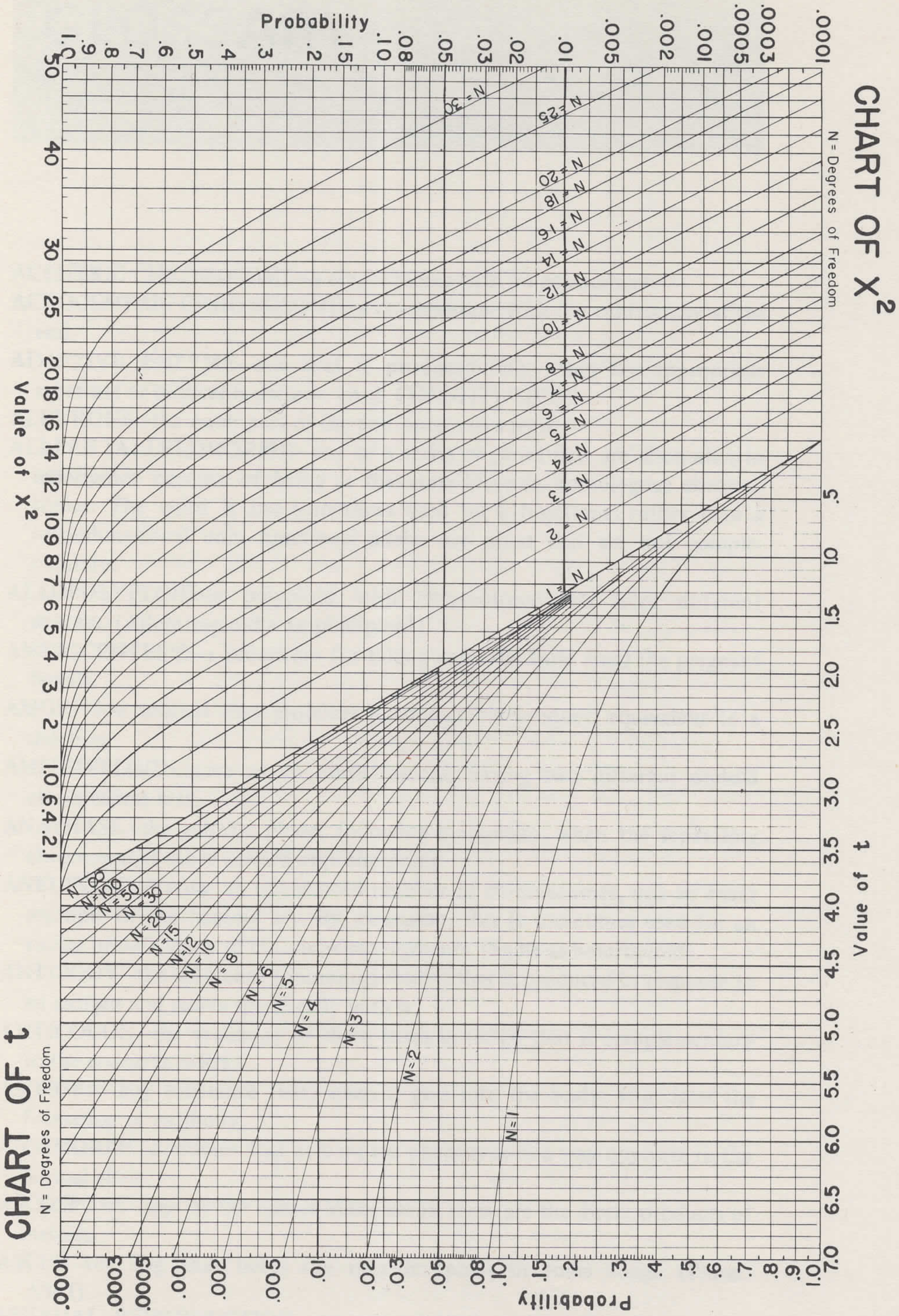


CHART OF X²
N = Degrees of Freedom
Probability
1.0 .9 .8 .7 .6 .5 .4 .3 .2 .15 .10 .08 .05 .03 .02 .01 .005 .002 .001 .0005 .0003 .0001
Value of X²
50 40 30 25 20 18 16 14 12 10 9 8 7 6 5 4 3 2 1.0 .6 .4 .2 .1
N = 30
N = 25
N = 20
N = 18
N = 16
N = 14
N = 12
N = 10
N = 8
N = 7
N = 6
N = 5
N = 4
N = 3
N = 2
N = 1
CHART OF t
N = Degrees of Freedom
Value of t
.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0
Probability
.0001 .0003 .0005 .001 .002 .005 .01 .02 .03 .05 .08 .10 .15 .2 .3 .4 .5 .6 .7 .8 .9 1.0
N = ∞
N = 100
N = 50
N = 30
N = 20
N = 15
N = 12
N = 10
N = 8
N = 6
N = 5
N = 4
N = 3
N = 2
N = 1

GLOSSARY

ACENTRIC CHROMOSOME–a chromosome with no centromere.

ACROCENTRIC CHROMOSOME–a chromosome with the centromere at the end.

ADAPTIVE ENZYME–one that is produced only when the appropriate substrate or inducer is present. (Ant. CONSTITUTIVE)

ALEURONE–the outer part of the endosperm in a seed.

ALLELE (ALLELOMORPH)–one of a series of genes that are alternative in inheritance because of being at the same locus in homologous chromosomes. The word is also sometimes used for a functional rather than a recombinational unit–functional alleles are genes that are not complementary.

ALLOPOLYPLOID–a polyploid with chromosome sets from different sources. (Allotetraploid = amphidiploid)

AMNIOCENTESIS–a procedure for taking amniotic fluid from the pregnant female.

AMORPH–a mutant that produces no detectable product; equivalent to a deletion.

AMPHIDIPLOID–same as an allotetraploid, having two different diploid chromosome sets.

ANAPHASE–the mitotic phase, following metaphase, when the separating chromosomes move toward opposite poles.

ANEUPLOID–having an unbalanced number of chromosomes, not an exact multiple of the haploid set. By extension, this is sometimes used for an excess or deficiency of chromosome segments. (Same as heteroploid)

ANTIBODY–the substance (immunoglobulin) that is produced in response to an antigen and reacts specifically with it.

ANTICODON–the sequence of three bases in tRNA that is complementary to the codon in mRNA.

ANTIGEN–any substance that, when it gets into the body, stimulates the formation of antibody.

ANTIMORPH–a mutant that acts on the phenotype in a way opposite to the normal allele.

ASCOSPORE–one of the spores in the ascus; these are the direct products of meiosis.

ASCUS–the bag that holds the meiotic spores in some fungi. (Plural: ASCI)

ASEXUAL REPRODUCTION–any reproductive process that does not involve meiosis and fertilization.

ASSORTATIVE MATING–mating of individuals that are more similar phenotypically than if mating were random.

AUTOGAMY–a process, such as occurs in Paramecium, where identical haploid nuclei fuse to form a completely homozygous diploid nucleus.

AUTONOMOUS–developing or behaving in accordance with its own genotype and not influenced by surrounding tissues.

AUTOPOLYPLOID–a polyploid in which the chromosomes all come from the same species. (Ant. ALLOPOLYPLOID)

AUTOSOME–a chromosome other than the sex chromosomes.

AUTOTROPHIC–having minimal nutritional requirements; not requiring complex organic compounds.

AUTOZYGOUS–having a pair of alleles that are descended from the same gene in an ancestor common to the mother and father. Autozygous means that the two alleles have the same historical origin; homozygous, that they are indistinguishable, but not necessarily derived from the same ancestral gene. (Ant. ALLOZYGOUS)

AUXOTROPH–requiring a particular nutrient, such as a vitamin or amino acid, or a combination of these, for growth.

BACKCROSS–a mating in which one member is homozygous for one or more genes. If the homozygous genes are recessive this is called a TESTCROSS.

BACTERIOPHAGE–a bacteria-infecting virus.

BALANCED LETHALS–two recessive lethals on homologous chromosomes that are both perpetuated in heterozygous condition because the two homozygous types are lethal. This requires that the two loci be very closely linked or that crossovers be suppressed by a chromosome rearrangement.

BALANCED POLYMORPHISM–two or more genotypes or phenotypes maintained in a population by balanced selective forces.

BARR BODY–(See SEX CHROMATIN BODY)

BIVALENT–two synapsed chromosomes, comprising four chromatids, ordinarily seen at the diplotene stage of meiosis.

C-MITOSIS–the chromosome behavior following treatment with colchicine and similarly-acting substances; the chromatids remain attached at the centromere and the nuclear division is not completed.

CENTROMERE–the chromosome region to which the spindle fiber attaches. (Syn. KINETOCHORE, PRIMARY CONSTRICTION, SPINDLE ATTACHMENT)

CENTROSOME–the self-replicating body in the cell to which the spindle fibers attach.

CHIASMA–the X-shaped figure seen at diplotene following a crossover. (Plural: CHIASMATA)

CHIMERA–an organism with a mixture of cells or tissues of more than one genotype. (See MOSAIC)

CHLOROPLASTS–chlorophyll-containing cytoplasmic particles in which photosynthesis takes place.

CHROMATID–one of the two daughters into which the chromosome divides, and which are still attached to the same centromere.

CHROMATID INTERFERENCE–nonrandom frequencies of 2-, 3-, and 4-strand double crossovers.

CHROMATIN–the substance of the chromosomes. (See EUCHROMATIN, HETEROCHROMATIN)

CHROMOMERE–one of the periodic enlargements along the chromosome thread.

CHROMONEMA–the chromosome thread. (Plural: CHROMONEMATA)

CHROMOSOME–the nuclear body, visible with the microscope, that carries the genes. By extension, the bacterial genome, a DNA thread, is sometimes called a chromosome even though it does not have the same structure as the chromosome in higher animals and plants.

CIS PHASE–the linkage phase where the mutant genes are on the same chromosome. (Ant. TRANS PHASE)

CIS-TRANS POSITION EFFECT–the condition where the cis and trans gene arrangements give different phenotypes.

CISTRON–a unit of chromosome length such that within this region recessive mutants are noncomplementary. (See LOCUS)

CLONE–a group of cells that are derived by mitotic divisions from a single ancestral cell.

CO-DOMINANCE–the state where each gene produces its characteristic effect in the heterozygote, as in the AB blood group genotype which produces both A and B antigens.

CODON–a unit of three nucleotides that form the code for a single amino acid.

COEFFICIENT OF INBREEDING–a measure of the proportion by which the heterozygosity is reduced by consanguinity; the probability of autozygosity.

COEFFICIENT OF KINSHIP–Same as coefficient of consanguinity. The probability that two alleles, one chosen at random from each of two individuals, are identical by descent. The value is the same as the coefficient of inbreeding of a progeny of these two individuals.

COINCIDENCE–the ratio of the observed double crossovers to the number expected if crossovers occurred at random; used as a measure of interference (q.v.).

COMPLEMENTATION–the production of the normal phenotype by two mutants when they are in the same cell. Mutant genes that do this are said to be COMPLEMENTARY.

CONCORDANT–alike with respect to some trait. (Ant. DISCORDANT) (Noun: CONCORDANCE)

CONSANGUINEOUS–having one or more ancestors in common; related.

CONSTITUTIVE ENZYME–an enzyme that is produced in the same amount whether or not its substrate or inducer is present. (Ant. INDUCIBLE)

COUPLING–the linkage phase where two dominant alleles are on one chromosome and the two corresponding recessives on the homologous chromosome. (Ant. REPULSION)

CROSSING-OVER–the exchange of segments between homologous chromatids, usually at meiosis.

CYTOPLASM–the material of the cell except for the nucleus.

CYTOPLASMIC INHERITANCE–transmission of hereditary characters by elements in the cytoplasm.

DEGENERATE–having two or more codes for the same amino acid.

DELETION, DEFICIENCY–the absence of a chromosome segment.

DICENTRIC–having two centromeres.

DIPLOTENE–the meiotic prophase stage when the four strands of the tetrad are clearly visible and chiasmata can be seen.

DISJUNCTION–regular separation and movement to opposite poles of chromosomes during nuclear division.

DIZYGOTIC, DIZYGOUS–arising from two fertilized eggs. (Syn. FRATERNAL)

DNA–deoxyribonucleic acid, the genetic material (except in some viruses).

DOMINANT–expressed when present in a single dose. Strictly, if Aa and AA are indistinguishable, A is said to be dominant to a. Often, especially in human genetics, any gene that produces an effect in the heterozygous state is said to be dominant. (Ant. RECESSIVE)

DOSAGE COMPENSATION–the regulation of the phenotype so that it is the same even though the number of causative genes varies, as with mutant genes on the X chromosome which produce the same phenotype in males and females.

DOUBLECROSS–in plant breeding, the mating of two different F_1 hybrids, each made by crossing two inbred lines. (Syn. FOUR-WAY CROSS)

DRIFT–see GENETIC DRIFT

DUPLICATION–an extra segment of a chromosome, sometimes free but more often attached to or inserted in a chromosome.

DYSGENIC–harmful to the genetic makeup of future generations.

ELECTROPHORESIS–migration of particles in an electric field, used particularly to separate related proteins.

ENDONUCLEASE–an enzyme that cuts a nucleic acid strand (in the interior, not at the end). (See EXONUCLEASE)

ENDOSPERM–the triploid nutritive part of a grain seed (e.g., maize).

EPISOME–a genetic element that may either be integrated into the genome or exist as a cytoplasmic factor.

EPISTASIS–the masking of the expression of a gene by a gene at another nonallelic locus. If the phenotypic effect of *A* and *a* cannot be distinguished in the presence of *b*, then *b* is said to be epistatic to the *A* genes. By extension, any deviation from additive effects of nonhomologous genes is also called epistasis, especially in discussion of population and quantitative genetics. (Syn. INTERACTION)

EQUATIONAL DIVISION–separation of sister chromatids. (Ant. REDUCTION DIVISION)

EUCHROMATIN–the regular staining material of most of the chromosome that contains most of the genes.

EUGENICS–"the study of agencies under social control, that may improve or impair the racial [meaning hereditary] qualities of future generations." (Galton)

EUKARYOTE–an organism with a nucleus and with chromosomes containing both DNA and protein. (See PROKARYOTE)

EUPLOID–having a balanced number of chromosomes, that is an exact multiple of the haploid set. (Syn. ORTHOPLOID; Ant. ANEUPLOID)

EUTHENICS–improvement of the human phenotype by environmental means.

EXONUCLEASE–an enzyme that removes bases from the end of a nucleic acid strand.

EXPRESSIVITY–the degree or quality of the phenotypic expression of a gene or group of genes.

F FACTOR–the male sex determining episome in bacteria. F^+ is male, F^- is female. (See also HFR)

FEEDBACK INHIBITION–the inhibition of enzyme activity by the product of a later stage in the sequence of reactions.

FILIAL REGRESSION–(See REGRESSION)

FITNESS–a measure of the capacity to survive and reproduce, usually measured relative to other genotypes.

FOUNDER PRINCIPLE–the divergence of a population from others because of the chance genetic makeup of a very small number of individuals from which it is descended.

FRAMESHIFT MUTATION–a mutation caused by deletion or duplication of one or a few nucleotides.

FRATERNAL TWINS–dizygotic twins (q.v.).

FREEMARTIN–an intersex-like, sterile female calf born twin to a male.

G_1–the interphase stage before DNA doubling.

G_2–the interphase stage after DNA doubling.

GAMETE–the male or female reproductive cell; sperm or ovum.

GAMETOPHYTE–the haploid part of the plant life cycle. (See SPOROPHYTE)

GENE–in classical genetics, the unit factor in inheritance. In modern usage it is equated to the functional unit responsible for synthesis of one polypeptide, or (where a complementation test is employed) a cistron.

GENETIC DRIFT–fluctuation of gene frequencies brought about by chance; better called Random Genetic Drift to distinguish it from steady processes.

GENETIC LOAD–the decrease of fitness, or other attribute of interest, in a population as a result of deleterious genes.

GENOME–a haploid set of chromosomes and the corresponding set of genes.

GENOTYPE–the genetic constitution of an organism.

GYNANDROMORPH–an individual with both male and female tissues; a sex-mosaic. (Cp. INTERSEX, HERMAPHRODITE)

HAPLOID–having a single set of chromosomes, as a gamete (See DIPLOID, POLYPLOID). Sometimes this is used to refer to the gametic number even if more than one set is present, as the gamete of a tetraploid. (See MONOPLOID)

HARDY-WEINBERG PRINCIPLE–the relationship giving the genotype frequencies in terms of gene frequencies in a randomly mating population.

HEMIZYGOUS–having only one representative of the gene or genes under discussion, as X-chromosome genes in the male. (See HOMOZYGOUS, HETEROZYGOUS)

HERMAPHRODITE–an individual with both testis and ovary, or with both male and female reproductive structures. (See INTERSEX, GYNANDROMORPH)

HETEROCHROMATIN–a part of the chromosome material that stains differently from the rest of the chromosomes, especially in some stages of the division cycle. (Ant. EUCHROMATIN)

HETEROGAMETIC–able to produce more than one kind of gamete as regards chromosome type, as the male which produces both X- and Y-bearing sperm. (Ant. HOMOGAMETIC)

HETEROGRAFT—a graft of tissue between individuals of different species.

HETEROKARYON—a cell with nuclei of more than one type. (Adj. HETEROKARYOTIC)

HETEROPLOID—(See ANEUPLOID)

HETEROSIS—hybrid vigor; the increase in size, performance, fitness, etc., that is usually associated with heterozygosity.

HETEROTROPHIC—requiring complex nutrients, such as amino acids and vitamins. (Ant. AUTOTROPHIC)

HETEROZYGOUS—having two different alleles at the locus or loci under discussion. (Ant. HOMOZYGOUS) (Noun: HETEROZYGOTE)

HFR—High frequency recombination; a strain of bacteria in which the F episome is integrated into the genome.

HOLANDRIC—found only in males, such as traits inherited through genes on the Y chromosome. (Ant. HOLOGYNIC)

HOMOGAMETIC—able to produce only one kind of gamete as regards chromosome type, as the female which produces only eggs with an X chromosome. (Ant. HETEROGAMETIC)

HOMOGRAFT—a graft of tissue between individuals of the same species.

HOMOLOGOUS CHROMOSOMES—the two members of a chromosome pair, one derived from each parent, that normally synapse at meiosis and have corresponding gene loci.

HOMOZYGOUS—having identical or indistinguishable alleles at the locus or loci under discussion. (Ant. HETEROZYGOUS) (Noun: HOMOZYGOTE)

HYPOMORPH—a mutant that acts qualitatively similar to the normal gene but less effectively.

HYPOSTATIC—suppressed by a gene at another locus. (Ant. EPISTATIC) (Noun: HYPOSTASIS)

IDENTICAL TWINS—monozygotic twins.

IMMUNOGLOBULIN—antibody globulin.

INBREEDING—mating of individuals more closely related than if they were chosen at random from the population.

INDEPENDENT ASSORTMENT—the failure of one pair of genes to be associated with the behavior of another pair during meiosis as when they are on nonhomologous chromosomes. Mendel's Second Law.

INDUCER—a substance that causes an increase in the amount of an adaptive enzyme.

INTERCROSS—a cross between two individuals each heterozygous for the locus or loci under discussion. (Cp. BACKCROSS, TESTCROSS)

INTERFERENCE—the tendency for one crossover to reduce to below the random expectation the frequency of others in its vicinity. (See COINCIDENCE)

INTERPHASE—the "resting" stage of mitosis when the nucleus is not dividing. (See G_1, G_2, and S stages)

INTERSEX—an individual intermediate between male and female. (Cp. HERMAPHRODITE, GYNANDROMORPH)

INVERSION—a chromosome rearrangement in which a segment is reversed.

ISOALLELES—alleles whose effects are indistinguishable except by special techniques.

ISOCHROMOSOME—a chromosome with two identical arms.

ISOGENIC—having identical genotypes.

ISOLATING MECHANISMS–mechanisms that prevent gene exchange between different species.

KARYOTYPE–the number, sizes, and morphology of the chromosomes of an individual.

KINETOCHORE–centromere (q.v.).

LEPTOTENE–the first stage in meiotic prophase when the chromosomes are thread-like and have not yet synapsed.

LETHAL EQUIVALENT–a group of genes that, if dispersed in different individuals, cause on the average one death.

LIGASE–an enzyme that joins two segments of nucleic acid.

LOCUS–the position of a gene on the chromosome or in the linkage map. Also used as a synonym for the gene, defined as a functional unit. If the experimental test is complementation, this is the same as a CISTRON.

LYSOGENIC–historically, capable of inducing lysis. Now applied to bacteria that carry a prophage.

MAPPING FUNCTION–a formula relating the chromosome map distance between two loci to the proportion of recombination between them.

MEGASPORE–the female cells produced by meiosis in a plant and which gives rise to the gametophyte.

MEIOSIS–the two nuclear divisions that lead to the formation of gametes, during which the chromosome number is halved.

MESSENGER RNA, or mRNA–RNA of high molecular weight that carries genetic information from the gene to the cytoplasm.

METACENTRIC CHROMOSOME–one having the centromere somewhere near the middle.

METAFEMALE–an individual with three X chromosomes. (Formerly SUPERFEMALE)

METAPHASE–the mitotic phase, following prophase, in which the chromosomes lie along the equatorial plane.

MICROSPORE–the male haploid cell produced by meiosis in a plant; this may function as a gamete or after a few mitotic divisions give rise to the gamete.

MISSENSE MUTATION–one that changes a codon so that it produces a different amino acid.

MITOSIS–nuclear division in which the daughter chromosomes pass to opposite poles. (See PROPHASE, METAPHASE, ANAPHASE, TELOPHASE, INTERPHASE)

MODIFIER–a gene which modifies the expression of another gene or genes.

MODULATION–control of gene action at the translation level.

MONOPLOID–having a single set of chromosomes. (See HAPLOID)

MONOSOMY–the occurrence of only one representative of a chromosome in an otherwise diploid cell.

MONOZYGOTIC, MONOZYGOUS–arising from the same zygote, as one-egg twins. (Syn. IDENTICAL) (Ant. DIZYGOTIC)

MORGAN–a unit of recombination, equal to 100 map units (centimorgans).

MOSAIC–an individual containing cells of two or more genotypes (Syn. CHIMERA).

MULTIPLE FACTORS–(See POLYGENIC)

MUTAGEN–an agent that increases the mutation rate.

MUTANT–a mutated gene or an individual showing the effects of the mutated gene.

MUTATION–a change of a gene from one allelic form to another. Sometimes used more broadly to include chromosomal aberrations.

MUTON–the smallest unit of chromosome that can undergo mutation (the same as a nucleotide).

NONDISJUNCTION–a failure of chromosomes to go to opposite poles during cell division, leading to unequal numbers of chromosomes in the daughter cells.

NONSENSE MUTATION–one that changes an amino acid codon to a terminating codon.

NUCLEIC ACID–DNA or RNA (q.v.).

NUCLEOLUS–a body in the nucleus, associated with a particular region of the chromosome, and which is involved with ribosomal RNA.

NUCLEOSIDE–a purine or pyrimidine together with a 5-carbon sugar.

NUCLEOTIDE–a nucleoside with phosphate; the unit building block for nucleic acid.

NUCLEUS–the inner part of the cell, containing the chromosomes.

NULLISOMIC–having one chromosome missing entirely.

OPERATOR GENE–a gene that is responsible for regulating the function of a group of genes closely linked to it.

OPERON–a group of linked genes that are under the control of a single operator gene.

OVERDOMINANCE–the condition where the heterozygote is more extreme for some trait than either homozygote.

OVUM–the female gamete, or unfertilized egg.

PACHYTENE–the stage in meiotic prophase when synapsis is complete.

PANMIXIS–random mating.

PARACENTRIC–to one side of the centromere. A paracentric inversion does not include the centromere in the inverted region.

PARTHENOGENESIS–development of an unfertilized egg.

PENETRANCE–the proportion among those having the requisite genotype that show the characteristic phenotype.

PERICENTRIC–around the chromosome. A pericentric inversion includes the centromere in the inverted region.

PHENOCOPY–an environmentally-produced change in phenotype that mimics a trait known to be caused by a genetic change.

PHENOTYPE–the character produced by the genes, with the environment; in contrast to the GENOTYPE, or genetic makeup.

PLEIOTROPY–the production of several, superficially unrelated, effects by a single gene.

POLYGENIC–influenced by many genes. (Syn. MULTIFACTORIAL)

POLYMERASE–an enzyme that catalyzes the building up of a polymer from its unit parts. DNA polymerase catalyzes the the synthesis of DNA.

POLYMORPHIC–having two or more distinct phenotypes both maintained in the population. Usually this implies that the types are maintained by some mechanism other than recurrent mutation. (Noun: POLYMORPHISM)

POLYPEPTIDE–a string of amino acids joined together by peptide linkages.

POLYPLOIDY–the occurrence of more than two haploid sets of chromosomes in a cell. Levels of polyploidy are designated as TRIPLOIDY, TETRAPLOIDY, PENTAPLOIDY, etc. (Adj. POLYPLOID)

POLYSOMY–the occurrence of more than two representatives of a chromosome in an otherwise diploid cell. (Adj. POLYSOMIC)

POLYTENE–many stranded, as the salivary gland chromosomes of diptera.

POSITION EFFECT–the dependence of the phenotype on the arrangement of the genes on the chromosome. (See CIS-TRANS POSITION EFFECT, V-TYPE POSITION EFFECT)

PRIMARY OOCYTE–the cell that after two meiotic divisions produces the ovum.

PRIMARY SPERMATOCYTE–the cell that, after two meiotic divisions, produces four spermatozoa.

PROBAND–an individual with a hereditary trait who is discovered independently rather than through his relationship to another affected individual. (See PROPOSITUS)

PROMOTER–the region of the DNA to which RNA polymerase binds and where RNA synthesis begins.

PROPHAGE or PROVIRUS–a virus that is integrated into the chromosome.

PROPHASE–the first phase of mitosis, following interphase, in which the chromosomes appear to contract and thicken.

PROPOSITUS–the index case of a disease or trait through which a family is discovered. (See also PROBAND)

PROTOTROPH–an organism that is able to grow without a specific nutrient or group of nutrients. (Ant. AUXOTROPH)

PSEUDODOMINANCE–the expression of a recessive gene when the dominant allele is missing, as with a deletion or hemizygosity.

R–the ROENTGEN, a unit of ionizing radiation. In protoplasm it is about two ionizations per cubic micron.

RAD–a unit of radiation measured by the amount of energy absorbed (.01 joules per kg).

RECOMBINATION–reassortment of genes so that a gamete contains genes of both paternal and maternal origin. If the genes are linked, this occurs by crossing-over.

RECON–the smallest region between which but not within which recombination occurs. (Syn. SITE)

REDUCTION DIVISION–separation of homologous chromatids. (Ant. EQUATION DIVISION)

REGRESSION–(1) the tendency for the progeny, or other relatives, of extreme individuals to be nearer the population average; (2) in statistics, the amount by which the independent variable changes per unit change in the dependent variable.

REGULATOR GENE–a gene that influences the amount of product (such as a polypeptide) produced by another gene.

REM–a unit of radiation corrected for biological effectiveness to equal one rad of X rays.

REPRESSION–prevention of the synthesis of an enzyme.

REPRESSOR–the substance, produced by a regulator gene, that represses enzyme formation.

REPULSION–in a double heterozygote, the linkage phase where the two dominant alleles are on opposite chromosomes. (Ant. COUPLING)

REVERSE TRANSCRIPTASE–an enzyme that catalyzes the formation of DNA from an RNA template.

RIBOSOME–cytoplasmic particle concerned with protein synthesis.

RNA–ribonucleic acid. (See MESSENGER RNA, TRANSFER RNA)

ROENTGEN–a unit or radiation measured by the number of ionizations (in tissue it is about 2 ionizations per cubic micron)

S-Stage–the interphase stage, between G_1 and G_2, when the DNA synthesis takes place.

SATELLITE–a tiny terminal segment of a chromosome separated from the rest of the chromosome by a constriction.

SECONDARY OOCYTE–the cell produced, along with a polar body, by the first female meiotic division.

SECONDARY SPERMATOCYTE–one of the two cells produced by the first male meiotic division.

SEGREGATION–separation of homologous chromosomes or genes into separate gametes, as occurs regularly during meiosis. Mendel's First Law.

SELECTION–differential survival or fertility of one type compared with others. NATURAL SELECTION occurs in nature; ARTIFICIAL SELECTION is practiced by the animal or plant breeder.

SELECTION INDEX–a value assigned to an individual animal or plant that indicates his breeding worth; it combines in an optimum way information from various measurements or observations on the individual and his near relatives.

SEX CHROMATIN BODY–a stainable body derived from an X chromosome and found in the interphase nucleus of cells which have two or more X chromosomes.

SEX-CHROMOSOME–the X or Y chromosome.

SEX-DUCTION–transduction by means of the F factor.

SEX-LIMITED–restricted to one sex.

SEX-LINKED–located on the X or Y chromosome; usually used more specifically to refer to genes on the X. Also used for traits caused by such genes. (See X-LINKED, Y-LINKED)

SIBS, SIBLINGS–brothers, or sisters, or brothers and sisters.

SIGMA–(1) a cytoplasmic factor for carbon dioxide sensitivity in Drosophila; (2) the part of RNA polymerase that recognizes the promoter, and thus initiates transcription.

SITE–recon.

SOLUBLE RNA, sRNA–(See TRANSFER RNA)

SOMATIC CELLS–all the body cells except the germ cells.

SPERMATOCYTE–a cell that gives rise to sperm cells. (See PRIMARY SPERMATOCYTE and SECONDARY SPERMATOCYTE)

SPERMATOZOON, SPERM–the male gamete.

SPINDLE-FIBER ATTACHMENT–see CENTROMERE.

SPOROPHYTE–the diploid part of a plant life cycle.

STRUCTURAL GENE–a gene that determines the amino acid sequence of a polypeptide.

SUPPRESSOR–a gene which nullifies the effect of another, usually mutant gene.

SYNAPSIS–the pairing of homologous chromosomes, such as occurs in the meiotic prophase.

SYNDROME—a group of symptoms that occur together in a particular disease or abnormality.

TELOPHASE—the mitotic phase, following anaphase, when the two groups of chromosomes begin to return to the interphase condition.

TEMPERATE PHAGE—a bacterial virus that, rather than lysing the cell, may be incorporated into the genetic makeup of the cell.

TEMPLATE—a pattern or mold.

TERMINALIZATION—moving of the chiasma away from the centromere in the late meiotic prophase.

TESTCROSS—a cross to an individual homozygous for one or more recessive genes. (See BACKCROSS)

TETRAPLOID—having four haploid sets of chromosomes.

TRANS PHASE—the linkage phase where the two mutant genes are on opposite chromosomes. (Ant. CIS PHASE)

TRANSCRIPTION—formation of RNA from a DNA template.

TRANSDUCTION—the carrying of DNA from one bacterium to another by a virus.

TRANSFER RNA, tRNA—low molecular weight RNA that attaches to an amino acid and guides it to the correct position in protein synthesis. (Syn. SOLUBLE RNA)

TRANSFORMATION—the transfer and incorporation of DNA from one cell to another (as in some bacteria).

TRANSITION—mutation of a purine into another purine, or pyrimidine into another pyrimidine.

TRANSLATION—formation of a polypeptide in the ribosome in a sequence determined by the messenger RNA.

TRANSLOCATION—the attachment of a chromosome segment to another chromosome; the reciprocal exchange of parts between nonhomologous chromosomes, or more rarely between homologous chromosomes broken at different points.

TRANSVERSION—mutation of a purine into a pyrimidine, or vice versa.

TRIPLOID—having three haploid sets of chromosomes.

TRISOMIC—having three representatives of one chromosome.

V-TYPE POSITION EFFECT—an altered phenotype caused by a chromosome rearrangement such that a gene is moved from its normal euchromatic position to the heterochromatin.

X-LINKED—located on the X chromosome; also used for traits caused by such genes.

Y CHROMOSOME—the sex chromosome that is normally found singly in the male and not in the female.

Y-LINKED—located on the Y chromosome.

ZYGOTE—a cell formed by the fusion of male and female gametes.

ZYGOTENE—the stage in the first meiotic prophase when the homologous chromosomes have started to synapse.

ANSWERS TO PROBLEMS

CHAPTER 1

1. 1/2.
2. *B/b*; i.e., he is heterozygous.
3. Brown-eyed parents with a blue-eyed child, since both parents could be heterozygous.
4. No. More likely homozygous. Yes, for in this case the man must be heterozygous (unless there was mutation, which is very unlikely).
5. 1/4 solid colored, 1/2 roan, 1/4 white. 1/2 roan, 1/2 white. All roan.
6. Mate with a single-combed hen. If any single-combed progeny occur then he is heterozygous. If none appear, among a large number of rose-combed, he is probably homozygous.
7. The dominant factor, since removal of all showing the recessive phenotype would still leave recessive genes among heterozygous carriers.
8. 1/2 yellow, 1/2 gray. 2/3 yellow and 1/3 gray (among survivors). The first cross should produce larger litters, since 1/4 of the second cross would die as embryos.
9. 3/16. 1/16.
10. 1/4 brown-eyed taster, 1/4 blue-eyed taster, 1/4 brown-eyed non-taster, 1/4 blue-eyed non-taster.
11. 1/2. None
12. 9/16 black, 3/16 red, 3/16 liver, 1/16 lemon, since the black dog must have been doubly heterozygous.
13. 1/4.
14. 8.
15. 1/64.
16. 1/8.
17. *C/– l/l* and *c/c L/–*.
18. *c/c L/L* and *C/C l/l*. Both *c/c L/L* or both *C/C l/l*. Both *C/c L/l*, for example.

CHAPTER 2

1. In double-stranded DNA the number of thymine molecules should equal the number of adenine, and the number of guanine should equal the number of cytosine. In single-stranded DNA there is no necessary relationship among the bases.
2. ACTACATGGAATAT
   TGATGTACCTTATA

(The center piece must be both reversed and inverted to maintain the polarity of the sugar-phosphate sequence.)

3. 4^{100}. 2^{100}.
4. 1200 nucleotides.
5. Assuming that the two original strands were labeled but the new strands were not, two would be (half) labeled and two unlabeled.
6. If this happened during replication it would create an AC pair instead of the expected AT. At the next replication these would produce an AT and a GC pair. The GC would be a stable, new, mutant type.
7. About 4.5 million. About 1.5 mm.
8. About 2 meters in a diploid cell. Yes, it would go to the moon half a million times.
9. $4 \times 4 \times 4 = 64$.
10. $2 \times 2 \times 2 = 8$.

CHAPTER 3

1. 100. 400.
2. 100.
3. 46. 46. 46. 23. 23. 23.
4. Mother.
5. 8. 2^{23}, or about 8 million. (This assumes no exchange of parts between homologous chromosomes. See Chapter 8.)
6. Embryo started as a heterozygote, *B*/*b*. During development a cell whose descendants will form the blue eye had a mutation from *B* to *b*, or the *B*-containing chromosome got lost.
7. Assuming (correctly) that the extra chromosome goes randomly to one of the two cells at the first meiotic division, half normal and half mongoloid.

CHAPTER 4

1. Daughters normal; half the sons color-blind.
2. Cross-barred females with non-barred males, and the barred chicks will be male. (Don't forget that birds have XY females and XX males. Barred refers to the feather pattern, and has nothing to do with Barr bodies, in case this is a source of confusion.)
3. The Y chromosome.
4. 2/3 females, 1/3 males.
5. That YY females are of normal viability. Such females could produce only daughters, so the hypothesis could be tested this way. (Of course another way to test this hypothesis would be to look at the chromosomes, assuming that the X and Y can be told apart.)
6. Females: half black, half tortoise-shell; Males: half black, half yellow. You would not expect any tortoise-shell males, since only those with two X's could be heterozygous.
7. They are XXY and, like the corresponding human type, are sterile.
8. Mother *C*/*c* *B*/*b*, father *C*/*Y* *B*/*b*.

9. 1/4.
10. Males chartreuse, females brown.
11. A male bee.
12. I-1, *C/c*; I-2, *C*; II-1, unknown but probably *C/C* since the *c* gene is rare; II-2, *c*; II-3, *C/c*; II-4, *C*; III-1, *C/c*; III-2, *C/C* or *C/c*; III-3, *c*; III-4, *C*.
13. Nondisjunction (or chromosome loss) must have occurred in the father, since he had to transmit a sperm with neither an X or Y. Under other circumstances the child would have been color-blind.
14. Among the surviving genotypes, 1/3 *T/t* females, 1/3 *T* males, 1/3 *t* females.
15. Perhaps she carried a recessive lethal gene on one of her X chromosomes.
16. Transplant the left gonad into a previously castrated bird, hoping it will develop into an ovary. The remaining right gonad, perhaps with a little help from hormone treatments, may develop into a testis. If all this works, the two birds could be mated to produce the genetic equivalent of self-fertilization.
17. Females all normal eyed; half the males normal and half ivory.
18. 2.
19. One possibility is that not all genes on the X are inactivated. Another is that two X's are required for normal early embryonic development before the inactivation occurs.
20. Essentially zero. Since he can transmit only sperm with a double chromosome or none, his children would either be mongoloid or monosomic and die as embryos. There could be normal children only if he married someone of his own genotype.
21. Nondisjunction at the second meiotic division of the mother, who was heterozygous for color blindness. (Another explanation will be apparent after you read Chapter 8.)
22. (a) all female; (b) 4/5 female and 1/5 male among the 5/6 of the eggs which hatch; (c) all male among the half of the eggs which hatch.
23. Normal XX females, XXX females, XY males, and XXY males in equal numbers.
24. One consequence is that daughters would not express genes from the father, but both sons and daughters would show (half the time) genes from the paternal grandfather.
25. XY from nondisjunction in the first division of the father and YY from nondisjunction in the second division. These would lead to XXY and XYY sons.
26. The fertilized egg separated into two individuals at the 2-cell stage. The Y chromosome got lost from one of them; the other having the original XY composition.
27. Mate *ba/ba Ts/ts* with *ba/ba ts/ts.* The former would produce pollen and the latter seeds, and the progeny would be of two equally frequent types like the parents.

CHAPTER 5

1. 4,196/148,462, or about .03.
2. 11,632/148,462, or about .08 assuming you can't die of both.
3. a. 5/16; B. 1/64; c. 1/64; d. 1/32; e. 11/32.
4. a. 1/256; b. 81/256; c. 27/64.
5. 3/4.
6. a. $\frac{100!}{75!\ \ 25!}\left(\frac{3}{4}\right)^{75}\left(\frac{1}{4}\right)^{25}$, which if you work it out is about .09.
   b. The sum of the probabilities of 71, 72, 73, 74, 75, 76, 77, 78, and 79 green plants, which turns out to be about .70.
   c. $1 - .70 = .30$.
7. $\frac{7!}{3!\ \ 1!\ \ 1!\ \ 2!}\left(\frac{3}{8}\right)^{3}\left(\frac{3}{8}\right)\left(\frac{1}{8}\right)\left(\frac{1}{8}\right)^{2}$, which is 0.016.
8. $\frac{556!}{315!\ \ 101!\ \ 108!\ \ 32!}\left(\frac{9}{16}\right)^{315}\left(\frac{3}{16}\right)^{101}\left(\frac{3}{16}\right)^{108}\left(\frac{1}{16}\right)^{32}$
9. $\chi^2_3 = .47$; Prob. = .93.
10. $1/2^{23}$, or about 1 in 8.4 million
    $1/2^{46}$, or about 1 in 70 million million } assuming no crossing-over (Chapter 8).
11. Solve the equation $(3/4)^n = .01$. The answer is 16.
12. 13/16.
13. $(1/10{,}000)\ (1/4)\ (1/4) = 1/160{,}000$.
14. 7/8.
15. No. It will alter the sex ratio in different-sized sibships but not the overall sex ratio.
16. $7!/7^7 = .0061$.

CHAPTER 6

1. Autosomal dominant. 1/2. 0, except for mutation or other complication.
2. Y chromosome (unless it is only a coincidence that only males are affected). 1/2. 0.
3. Autosomal recessive. Very small. Very small. (The exact probabilities depend on how common the gene is in the population.
4. X-linked recessive. Very small. 1/8 (1/4 for sons, 0 for daughters).
5. Autosomal recessive. Very small. Very small.
6. 1/2.
7. 1/3.
8. 1/420. Yes, for the cousin may carry the same gene inherited from a common grandparent.

CHAPTER 7

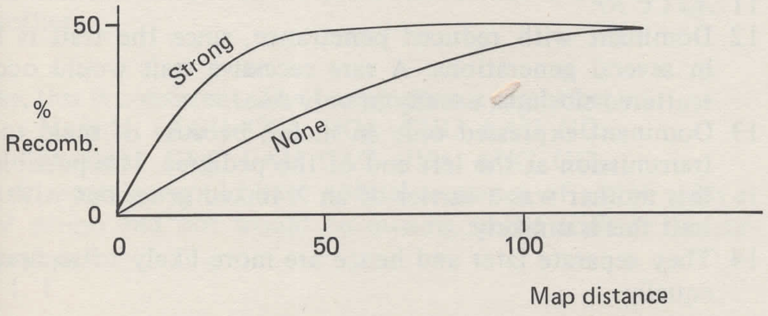
1. *A/A B/B*, *A/A b/b*, *a/a B/B*, and *a/a b/b*, each with probability 1/4.
2. C. For an explanation of A and D, see page 62.
3. The *l m* type wouldn't grow in minimal medium and hence wouldn't be detected.
4. *Aaa BBb*, *Aaa bbb*, *aaa BBb*, and *aaa bbb* in equal proportions.

5. Autogamy produces immediate, complete homozygosity. Self-fertilization makes half the loci homozygous, but complete homozygosity is approached only gradually.

CHAPTER 8

1. 21, or 22 if there are genes on the Y chromosome.
2. 25 Bar carnation, 25 normal, 475 Bar, 475 carnation.
3. *S* 5 *P* 3 *R* 10 *C*
4. *w* 10 *r* 20 *s*

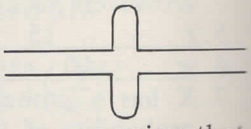
   (Note: if you correct for undetected doubles in region 2 by Kosambi's formula, the distance from *r* to *s* is about .21.) There is interference for the observed proportion of double crossovers in about half the number expected if the double crossovers occurred at random.
5. *r* 15 *z* 5 *s*
6. *w* 10 *r* 15 *z* 5 *s*
7. X has a probability of .1 of being a carrier, while Y has a probability of .9. Hence Y is much more likely to have affected sons.
8. *C* is in the middle.
9. 9/40 *AB*/*AB*, 1/40 *AB*/*Ab*, 1/40 *AB*/*aB*, 9/20 *AB*/*ab*, 1/40 *Ab*/*ab*, 1/40 *aB*/*ab*, 9/40 *ab*/*ab*.
10. The *A* and *T* loci are on different chromosomes, each close to its centromere.
11. The loci are close to each other, but far from the centromere.
12. 100 units.
13. No, since there is no upper limit to the length of the chromosome. If the length is greater than 100, this means that there is an average of more than one exchange per strand. (However, there is an upper limit of 1/2 for the proportion of recombination between two pairs of marker genes.)
14. No, centromere distance requires knowledge of the spore order. (However, if you have three loci on three chromosomes you can determine each centromere distance without ordered asci. Can you figure out how to do it?)
15. *a* 3 *b* 4 *c* 7 *l* 3 *d* 2 *e* 6 *f*
16. This could happen if the mother is heterozygous and there is an exchange between the color-blind locus and the centromere prior to nondisjunction. It could also be explained by nondisjunction at the second meiotic division without an exchange.
17.



18. Use males as the heterozygous parents, to take advantage of the fact that there is no crossing-over in Drosophila males.
19. 1/4 *a* +, 1/4 + *b*, 1/2 + +. (The amount of recombination in the female is irrelevant.)

CHAPTER 9

1. *N X A R C*, or the reverse.
2. *j k l a b c d e f g h i F*, or *i h g f e d c b a l k j F*.
3. One way would be to find progeny containing genes from each of 3 phages from a triply infected bacterium.
4. The trait would always be passed to only one of the two daughter cells at each cell division.
5. Only a single exchange is required to insert a circular region into a chromosome.
6. This permits the DNA to pair this way: This may provide a landmark for the enzyme to recognize.



7. One possibility is that they came from the same organism that produced the antibiotic, since this organism would have to be resistant to the antibiotic.

CHAPTER 10

1. .05.
2. *Bb*.
3. Transmission from father to son, when you can rule out transmission from the mother, would be inconsistent with X-linked inheritance. An exception to father-son transmission would rule out Y-linked inheritance. If you had a large amount of family data you could measure the correlation in incidence in fathers and sons; this should be zero for X-linked inheritance. On the other hand, with X-linked inheritance, there should be a correlation between incidence in a person and that in his maternal uncles and grandfather.
4. 5/8 normal, 3/8 sable.
5. (a) *Dd Ee* by *Dd Ee;* (b) *dd EE* by *DD ee,* for example.
6. 3/4.
7. In both sexes, 3/8 barred, 3/8 solid colored, 1/4 white.
8. (a) colored; (b) white; (c) 1/4 colored, 3/4 white; (d) 1/2 colored, 1/2 white.
9. 0.4.
10. Female.
11. *Aa CC Rr*.
12. Dominant with reduced penetrance, since the trait is found in several generations. A rare recessive trait would occur in scattered sibships, usually in only one.
13. Dominant expressed only in males, because of male to male transmission at the left end of the pedigree. It is possible that this mother was a carrier of an X-linked gene, but with a rare trait this is unlikely.
14. They separate later and hence are more likely to separate unequally.

15. Sibs.
16. For most traits both a proper genotype and a proper environment are essential. In Chapters 23 and 24 we consider a more meaningful question.
17. C. (B suggests that the cause is age dependent. For example if susceptibility to a disease is greatest at a certain age and a pair of twins is at this age during an epidemic, the probability of both being affected is increased.)
18. The proportion of opposite-sexed twins is the proportion of dizygotic times the proportion of opposite-sexed, given that they are dizygotic. The latter probability is 2P(1 - P), hence, from the formula at the bottom of page 40:

    Prop. opposite-sexed = Prop. dizygotic x 2P(1 - P)

    Rearrange this algebraically and you have the desired formula.
19. U.S. white 0.66; Japanese 0.38.
20. 1/4.
21. (a) All right; (b) 1/2 right, 1/2 left.
22. (a) *A– B– C–* have one phenotype (say they are pigmented), all others being the other phenotype (say, white). For example, the three dominant genes could be responsible for enzymes catalyzing three successive stages in formation of a pigment. (b) 1/8 pigmented, 7/8 white.
23. *Rr*.

CHAPTER 11

1. Y precedes X and the gene affects the step from Y to X.
2. $\xrightarrow{x}$ cysteine $\xrightarrow{w}$ cystathionine $\xrightarrow{r}$ homocysteine.
3. This argues that the damage is done by accumulated phenylalanine, not by absence of tyrosine.
4. Either of two recessive genes, when homozygous, leads to albinism. The man and woman were homozygous for different albino genes.
5. The second is much more likely, since ornithine precedes the other two in the urea cycle.
6. citrulline $\xrightarrow{2}$ arginosuccinic acid $\xrightarrow{1}$ arginine.
7. Mutants causing complete inactivation might well be lethal in a system as important as this.
8. Perhaps the proportion of cells where the mutant gene is on the inactivated X chromosome differs from one person to another.

CHAPTER 13

1. Yes, this is consistent. The base sequence might be:

    Normal: AGU CCA UCA CUU AAU GC–
    Mutant: GUC CAU CAC UUA AUG GC–
2. Short proteins would have normal sequence of amino acids at the amino end but would be missing amino acids at the carboxyl end.
3. 2. 1. 1.

4. Yes. For example,

$$\begin{tikzcd} & \substack{\text{CGA}\\\text{arg}} \arrow[dr] & \\ \substack{\text{GGA}\\\text{gly}} \arrow[ur] \arrow[dr] & & \substack{\text{GGA}\\\text{gly}} \\ & \substack{\text{GAA}\\\text{glu}} \arrow[ur] & \end{tikzcd}$$

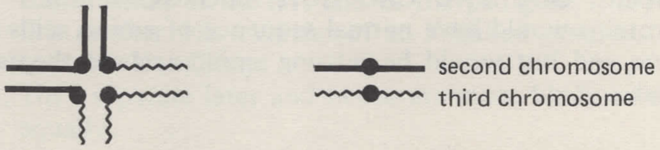
5. If it made this mistake regularly there would be a large number of misterminated polypeptides.
6. The mistranslation caused by streptomycin might produce a small number of molecules that have normal enzyme activity by happening to make a mistake that corrects the original mutant.
7. Normal protein synthesis is not required for chromosome donation; hence errors in translation or failure of protein synthesis do not matter for this process.
8. 5. This would provide 32 different sequences, more than enough.
9. GGG.
10. Most common: proline. Next: histidine, threonine. (Relative proportions would be 6:1:1.)
11. Leu, val, ileu, thr, lys, arg.
12. Minimum: 1 (UAI). Maximum: 4 (UAG, UAA, UAU, UAI).

CHAPTER 14

1. *White* and *cherry*.
2. C/c^{ch} and c/c.
3. 1/2 full color, 1/4 chinchilla, 1/4 white.
4. A complementation test identifies mutants as belonging to the same complementation group or to different groups, but it offers no information as to the sequence of the mutants within a group; they fail to complement regardless of their positions in the group. To determine the sequence requires recombination data.
5. Yes. Change the second or fourth so that they have the same number of facets.
6. s^1s^3 and s^1s^2 in equal proportions. None. s^1s^3, s^1s^4, s^2s^3, and s^2s^4 in equal proportions.
7. Both normal.
8. The *pearl* allele is efficient in competing for substrate but inefficient in converting it into a pigment. Hence it behaves like a hypomorph with inactive alleles, since it produces some pigment. With more active alleles, however, it takes substrate away from alleles that would more effectively convert it into pigment; in this way it acts as an antimorph.
9. (a) mutant; (b) normal; (c) normal; (d) mutant.

CHAPTER 15

1.



2. 1/2 normal, 1/2 *bw e*.

3. Because a recessive gene must be present 4 times for its phenotype to be expressed in a tetraploid as opposed to 2 times in a diploid.
4. (a) trisomic; (b) monoploid or haploid; (c) monosomic; (d) triploid; (e) autotetraploid; (f) nullosomic; (g) tetrasomic; (h) double trisomic.
5. Half the children would be expected to be mongoloid, assuming that the other parent had normal chromosomes.
6. a. By loss of the lower chromosome in the cell ancestral to the main parts.
   b. Autonomous: *s, e, c, g*. Nonautonomous: *v*. No information: *w, r, f*.
7. a. Two dicentric, two acentric; all unbalanced in gene content.
   b. All monocentric, but all unbalanced.

CHAPTER 16

1. Recessive mutant genes may be carried hidden many generations before becoming homozygous, so you don't know whether a homozygous recessive is due to a new mutation or simply to an old gene that has recently become homozygous. In the mouse and Drosophila one can use special experimental systems to measure recessive mutation rates.
2. If there were crossing-over, the lethal mutant might be removed from the + + chromosome and transferred to the w^aB chromosome. Hence there would be males with normal eyes and the culture would not be classified as having a lethal mutant.
3. Yes. Nitrous acid can produce changes both from AT to GC and the reverse.
4. UA, TA, CG, and CG.
5. No. Base substitution mutations would not be reverted by duplications or deletions.
6. Many are probably deletions.
7. Deficient repair enzymes is a likely explanation.
8. It might be a single-stranded DNA or RNA virus.
9. No. Nitrous acid produces transitions.

CHAPTER 17

1. One way would be this: Let σ stand for the cytoplasmic factor. Let an *aa bb* female mate with both *AA bb* and *aa BB* σ. Then, the finding of *Aa bb* σ progeny would show that the factor did not have to come from the sperm that contributed the *A b* chromosome. The fertilization of an egg by more than one sperm could be demonstrated by direct microscopic observation (though it probably wouldn't be easy).
2. A: *KK kappa;* B: *KK;* C: *kk.*
3. *Kk kappa* with killer phenotype.
4. If both transcribed strands are translated, then mutations should occur in pairs, one corresponding to each of the changed mRNA's. (This would surely create an evolutionary problem, since it would be very unlikely ever to have two new mutants both being beneficial.)

CHAPTER 18

1. (a). Both are produced with or without an inducer. (b) Same answer.
2. There may be several F′ particles in the same cell.
3. It might be a deletion, or it might be a promoter mutation.
4. Left: regulator; Middle: structural; Right: operator.

CHAPTER 19

1. 1/8.
2. Only the first is possible (except, of course, for errors, mutation, etc.).
3. A or AB. A or O. The principle is that the incoming cells must not be agglutinated by the serum of the recipient; antibodies in the incoming sera are diluted to the point of becoming ineffective in slow transfusions. Hence persons of group O can donate to any group.
4. Anti-d.
5. .45 x .15 = .07.
6. Assuming the existence of all six antisera (including anti-d): *CDE/cde, CDe/cdE, Cde/cDE,* and *CdE/cDe.*
7. Number 2 could be.
8. No decision, although the rarity of *CDE/cde* suggests that number 1 is not the father. The additional data confirm this by ruling out number 1.
9. 1. *CDe/cde;* 2. *CdE/cde;* 3. *cde/cde;* 4. *CDe/CdE;* 5. *cde/cde;* 6. *CdE/cde;* 7. *CDe/cde.*
10. Positive, positive, positive, negative.
11. Parents 1 – baby D; 2 – B; 3 – C; 4 – A.
12. (a) none; (b) all; (c) none; (d) all; (e) $1/2^5$ or 1/32; (f) $1/4^5$ or 1/1,024; (g) $(10/16)^5$; (h) $(3/4)^5$.
13. Very small. Very small. About 1/4. (1/16, very small, 1/4.)

CHAPTER 20

1. 1/10. 81/99 homozygous, 18/99 heterozygous.
2. 1/144, assuming only a single locus; with the correct, two-locus hypothesis, the proportion would be less as explained on page 165.
3. Roughly consistent. However, if you do a χ^2 test you will see that there is a small but statistically significant excess of roans.
4. 5/7. 4/49.
5. $\sqrt{.1436} = .378$. $\sqrt{.1902} = .436$. $2(.378)(.436) = .331$
6. .15. .0225. Before merger, .025.
7. 0.183.
8. Roughly, 0.87 + .035. (See page 237.)
9. (a) p. (b) $q^2/(1 + q)$.

CHAPTER 21

1. 1/12.
2. 1/64.
3. They are all 1/8. No genetic basis for social attitude.
4. 3/32.
5. 13/128.

7. Letting σ stand for the cytoplasmic factor and putting the maternal parent on the left for each mating:

M.S. $aa\ bb\ \sigma \times AA\ bb$ $\quad$ $aa\ BB \times aa\ BB$
M.S. $Aa\ bb\ \sigma \times aa\ BB$
M.F. $Aa\ Bb\ \sigma$ (M.S.: male sterile, M.F.: male fertile)

8. 85 bushels.
9. a. $2p_1p_4 + 2p_2p_3$. b. $2p_1p_2$. c. $p_1^2(15/16) + p_1(1/16)$.
10. a. $2(p_{CDE}p_{cde} + p_{CDe}p_{cDE} + p_{cDE}p_{Cde} + p_{CdE}p_{cDe})$.
    b. $p_{cde}^2(15/16) + p_{cde}(1/16)$.
11. 3/8.

CHAPTER 22

1. 1/289.
2. 8 generations. 20.
3. 184.5 pounds. This requires information from Chapter 23.
4. Cross the hybrid back to the susceptible variety; select resistant progeny and cross these to the susceptible variety; again select resistant progeny, and backcross again. Continue until almost all the genes, except those closely linked to the resistance gene, are from the susceptible (but otherwise desirable) variety.
5. Interpose one generation of self-fertilization after each cross and select a resistant progeny to backcross. Continue this each generation. (You could shorten the time by selfing only after several generations, making sure to grow enough plants to assure having at least one that is homozygous for the recessive factors.)
6. 2/10,000.
7. 1/10,000.
9. 10 per genome. The average number of homozygous recessives in a child whose parents were cousins would be 10/16, so about half of the conceptions would lead to lethal progeny. There is no such increase in miscarriage, stillbirth, or childhood death rate from cousin marriages. The most likely explanation is that most recessive mutations are not completely recessive and are eliminated from the population through minor deleterious heterozygous effects. At least this is the case with Drosophila.
10. The human reproductive rate is too low to permit very many such loci each independently causing a death rate as high as that of the S gene.
11. 20 generations, or about 5 centuries.

CHAPTER 23

1. Light.
2. Medium.
3. Dark.

4. In the first case the parents might be *AA bb;* in the second, *Aa Bb.*
5. All *A/a B/b,* size 2. 3/16 size 3, 10/16 size 2, 3/16 size 1.
6. 4 pairs, no dominance.
7. *A– B– cc dd* is 6 feet, *A– B– C– D–* and *aa bb cc dd* are 5 feet, *aa bb C– D–* is 4 feet, etc. Crossing two parents, *AA BB cc dd* and *aa bb CC DD,* would give the observed results.
8. The severely retarded are likely to be caused environmentally (e.g., birth accident or a virus infection during pregnancy) or by a single recessive factor, both of which have a low recurrence risk. The mildly retarded are likely to be at the low end of the normal scale; whether the reason is genetic or environmental, it is likely to be repeated for other children in the same sibship.

CHAPTER 24

1. .218 38%.
   .015 3%.
   .340 59%.
2. Yes, in a completely inbred line all differences are nongenetic.
3. Coat color in guinea pigs.
4. 0.3.
5. Means are all 100. Variances are 10.0, 6.8, and 3.2. The broad heritability is 0.68. The narrow heritability cannot be computed from these data. The variances are additive, so there are no interactions or covariances.

APPENDIX I

1. (a) $\chi^2_1 = 4$. Prob. about .05.

   (b) $\chi^2_1 = 0.4$. Prob. about 0.5.

   (c) $\chi^2_1 = 40$. Prob. less than 0.0001.

   Increasing the sample size by a factor x while keeping the same proportions increases χ^2 by a factor x. Increasing the sample size by a factor x while keeping the absolute differences constant decreases the value by a factor x. Can you prove this?
2. $\chi^2_3 = 2.67$, Prob. about .45. No reason to reject hypothesis.
3.

Comparison	χ^2	D. F.	Probability
Ee vs. *ee*	6.05	1	.014
Bb vs. *bb*	.80	1	.37
Independence	18.05	1	<.0001
Total	24.90	3	<.0001

   Note that the χ^2 values and the degrees of freedom are additive.
5. 3.8, 6.6. 18.3, 23.2. 1.96, 2.6. By taking the square roots of the χ^2 value. (This applies only to the case for infinite degrees of freedom for t and 1 degree of freedom for χ^2.)
6. 9.3; 19.3; 29.3. As the number of degrees of freedom increases, the value of chi-squared for P = 0.5 approaches the number of degrees of freedom.

n the second,

5 size 1.

dd are 5 feet,
AA BB cc dd
.

vironmentally
pregnancy) or
a low recur-
t the low end
ic or environ-
ildren in the

nongenetic.

2. The broad
not be com-
, so there are

ping the same
g the sample
e differences
u prove this?
hypothesis.

dom are ad-

are roots of
inite degrees

m increases,
e number of

APPENDIX II

1. $p = \dfrac{C + 2No \pm \sqrt{C[C + 4oN(1 - o)]}}{2(N + C)}$
2. .174 and .055.
3. .159 and .041.

APPENDIX III

1. (a) mean = 10; (b) variance = 16; (c) standard deviation = 4; (d) standard deviation of the mean = 1.15; (e) rough 95% confidence limits 7.5–12.5.
2. t = 3.44, D. F. = 85, prob. about .001. The two varieties are highly significantly different.

REFERENCES

GENERAL

Goodenough, U., and R. P. Levine. 1974. *Genetics.* Holt, Rinehart, and Winston. (Good for microbial and molecular genetics.)

Srb, A., R. Owen, and R. Edgar. 1965. *General Genetics.* Freeman. (For many years the standard, but now beginning to show its age. Still excellent for all but very recent material.)

Stern, C. 1973. *Principles of Human Genetics.* 3rd ed. W. H. Freeman. (More complete than Sutton, but less up to date.)

Strickberger, M. 1968. *Genetics.* Macmillan. (Unusually complete textbook, very useful for filling in details and for references.)

Sturtevant, A. H. 1965. *A History of Genetics.* Harper and Row. (Authoritative and very objective; written by one who contributed greatly to the history.)

Sturtevant, A. H., and G. Beadle. 1939. *An Introduction to Genetics.* Saunders. (A classic.)

Sutton, E. 1975. *An Introduction to Human Genetics.* Holt, Rinehart, and Winston. (Elementary, clear account of human genetics.)

Swanson, C. P., T. Merz, and W. J. Young. 1967. *Cytogenetics.* Prentice-Hall. (Clear and elementary.)

Watson, J. D. 1970. *Molecular Biology of the Gene.* Benjamin. (Interesting, informative, and easy to read. The subject matter is exactly what the title says, and it's written by one of the founders.)

Whitehouse, H. L. K. 1969. *Toward an Understanding of the Mechanism of Heredity.* St. Martin's. (Particularly good for cytogenetics.)

CHAPTER 1

Mendel, G. J. 1865. "Versuche uber Pflanzen-Hybriden." *Verh. Natur. Ver. Brunn.* 4. (Mendel's classic. An English translation is published by Harvard University Press and as a supplement to *Principles of Genetics* by Sinnott, Dunn, and Dobzhansky.)

Punnett, R. C. 1908. *Mendelism.* Macmillan. (Of historical interest as one of the earliest accounts. It is also the source of the poultry comb example.)

CHAPTER 2

Kornberg, A. 1974. *DNA Synthesis.* Freeman. (An up-to-date account of current knowledge by one who has contributed greatly to it.)

Watson, J. D., and F. H. Crick. 1953. "A structure for desoxyribose nucleic acids." *Nature* 171:737-738. (The classic.)

CHAPTER 3

Du Praw, E. J. 1970. *DNA and Chromosomes.* Holt, Rinehart and Winston.

Fawcett, D. W. 1966. *The Cell.* Saunders.

Sutton, W. S. 1903. "The chromosomes in heredity." *Biol. Bull.* 4:213-251. (An early argument for the chromosomes as the carriers of the genes.)

Wolfe, S. L. 1972. *Biology of the Cell*. Wadsworth.

CHAPTER 4

Bridges, C. B. 1916. "Non-disjunction as proof of the chromosome theory of heredity." *Genetics* 1:1-52, 107-163. (The clinching argument. What a way to start a new journal!)

Ohno, S. 1967. *Sex Chromosomes and Sex-linked Genes.* Springer.

CHAPTER 5

Feller, W. 1968. *An Introduction to Probability Theory and Its Applications.* 3rd ed. Wiley. (In a class by itself, but not easy reading.)

Mosimann, J. E. 1968. *Elementary Probability for the Biological Sciences.* Appleton-Century-Crofts. (Very elementary.)

CHAPTER 6

Fraser, F. C. 1974. "Genetic counseling." *Amer. J. Human Genet.* 26:636-659.

McKusick, V. A. 1975. *Mendelian Inheritance in Man.* 4th ed. Johns Hopkins Press. (Computerized output of known Mendelian traits, mostly diseases.)

CHAPTER 7

Beale, G. H. 1954. *The Genetics of Paramecium Aurelia.* Cambridge Univ. Press.

Fincham, J. R. S., and P. R. Day. 1963. *Fungal Genetics.* Blackwell.

Lederberg, J., and E. L. Tatum. 1946. "Gene recombination in E. coli." *Nature* 158:558.

Srb, A. M., R. D. Owen, and R. S. Edgar. 1965. *General Genetics.* Freeman. (Chapter 4 is the source of much of the material in this chapter.)

CHAPTER 8

Lima-de-Faria, A. 1969. *Handbook of Cytology*. (An enormous, multi-authored compendium.)

Lindsley, D. L., and E. H. Grell. 1968. *Genetic Variations of Drosophila Melanogaster.* Carnegie Inst. of Washington, Pub. No. 627. (A catalog of Drosophila mutants and their locations.)

Rhoades, M. M. 1950. "Meiosis in maize." *J. Heredity* 41:58-67. (A clear summary with excellent photographs.)

Ris, H., and D. F. Kubai, 1970. "Chromosome structure." *Ann. Rev. Genet.* 4:263-294.

Sturtevant, A. H. 1913. "The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association." *J. Exp. Zool.* 14:43-59. (The original paper on chromosome mapping by linkage analysis.)

CHAPTER 9

Campbell, A. 1969. *Episomes.* Harper and Row. (Brief, but excellent.)

Hayes, W. 1968. *The Genetics of Bacteria and their Viruses*. John Wiley. (Very complete.)

Jacob, F., and E. L. Wollman. 1961. *Sexuality and the Genetics of Bacteria.* Academic Press.

Stent, G. 1963. *Molecular Biology of Bacterial Viruses.* Freeman.

Susman, M. 1970. "General bacterial genetics." *Ann. Rev. Genet.* 4:135-176.

CHAPTER 10

Newman, H. H., F. N. Freeman, and K. J. Holzinger. 1937. *Twins: A Study of Heredity and Environment.* University of Chicago Press.

Shields, J. 1962. *Monozygotic Twins Brought Up Apart and Brought Up Together.* Oxford University Press. (Both books deal with studies of separated identical twins.)

CHAPTER 11

Garrod, A. E. 1909. *Inborn Errors of Metabolism.* Oxford University Press. (The beginning of biochemical genetics.)

Harris, H. 1975. *The Principles of Human Biochemical Genetics.* American Elsevier. (Very interesting and easy to read.)

Stanbury, J. B., J. B. Wyngaarden, and D. S. Fredrickson, 1972. *The Metabolic Basis of Inherited Disease.* McGraw-Hill. (Everything you could conceivably want to know about this subject.)

Wagner, R. P., and H. K. Mitchell. 1964. *Genetics and Metabolism.* John Wiley.

CHAPTER 12

Dayhoff, M. O. 1972. *Atlas of Protein Sequence and Structure.* Vol. 5. National Biomedical Research Foundation. (A compilation of sequence and other structural information. Very useful because of its completeness.)

Ingram, V. 1965. *The Biosynthesis of Macromolecules.* Benjamin.

Kornberg, A. 1974. *DNA Synthesis.* Freeman.

Mechanism of Protein Synthesis. 1969. Cold Spring Harbor Symposium, Vol. 34.

Nomura, M. 1970. "Bacterial ribosome." *Bact. Rev.* 34: 228-277.

Transcription of Genetic Material. 1970. Cold Spring Harbor Symposium, Vol. 35.

CHAPTER 13

The Genetic Code. 1966. Cold Spring Harbor Symposium, Vol. 31.

Khorana, H. G. 1967. "Polynucleotide synthesis and the genetic code." *The Harvey Lectures* 62:73-91.

CHAPTER 14

Benzer, S. 1959. "On the topology of the genetic fine structure." *Proc. Nat. Acad. Sci.* 45:1607-1620.

Benzer, S. 1961. "On the topography of the genetic fine structure." *Proc. Nat. Acad. Sci.* 47:403-415.

Yanofsky, C. 1967. "Structural relationships between gene and protein." *Ann. Rev. Genet.* 1:117-138.

CHAPTER 15

Little, T. M. 1958. "Gene segregation in autotetraploids." *Botan. Rev.* 24: 318-339.

Miller, D. A., and O. J. Miller. 1972. "Chromosome mapping in the mouse." *Science* 178:949-954.

Stebbins, G. L. 1950. *Variation and Evolution in Plants.* Columbia University Press.

Turpin, R., and J. LeJeune. 1969. *Human Afflictions and Chromosomal Aberrations.* Pergamon Press.

CHAPTER 16

Drake, J. W. 1970. *The Molecular Basis of Mutation.* Holden Day.

Fishbein, L., W. G. Flamm, and H. L. Falk. 1970. *Chemical Mutagens.* Academic Press.

Muller, H. J. 1927. "Artificial transmutation of the gene." *Science* 66:84-87. (The first demonstration of mutagenesis by radiation.)

CHAPTER 17

Jinks, J. L. 1964. *Extrachromosomal Inheritance*. Prentice Hall.

Poulson, D. F., and B. Sakaguchi. 1961. "Nature of "sex-ratio" agent in Drosophila." *Science* 133:1489-1490.

Preer, J. R. 1971. "Extrachromosomal inheritance: Hereditary symbionts, mitochondria, chloroplasts." *Ann. Rev. Genet.* 5:361-406.

Sager, R. 1972. *Cytoplasmic Genes and Organellas.* Academic Press.

Willetts, N. 1972. "The genetics of transmissible plasmids." *Ann. Rev. Genet.* 6:257-268.

CHAPTER 18

Brenner, M., and B. N. Ames. 1971. "The histidine operon and its regulation." In *Metabolic Regulation,* H. J. Vogel (ed.), Academic Press.

Brown, D. D., and I. B. Dawid. 1969. "Developmental genetics." *Ann. Rev. Genet.* 3:127-154.

Jacob, F., and J. Monod. 1961. "Genetic regulatory mechanisms in the synthesis of proteins." *J. Molec. Biol.* 3:318-356. (A paper famous for both content and presentation.)

Metzenberg, R. L. 1972. "Genetic regulatory systems in Neurospora." *Ann. Rev. Genet.* 6:111-133.

Reznikoff, W. S. 1972. "The operon revisited." *Ann. Rev. Genet.* 6:133-156.

CHAPTER 19

Fudenberg, H. H., J. R. L. Pink, D. C. Stites, and A. C. Wang. 1972. *Basic Immunogenetics.* Oxford University Press.

Hildeman, W. H. 1970. *Immunogenetics.* Holden Day.

Race, R. R., and R. Sanger. 1973. *Blood Groups in Man.* 6th ed. F. A. Blackwell.

Terasaki, P. 1970. *Histocompatibility Testing.* Munksgaard. (A collection of writings by many authors.)

CHAPTER 20-22

Cavalli-Sforza, L., and W. Bodmer. 1971. *The Genetics of Human Populations.* W. H. Freeman. (Emphasis on human populations, but also good for general theory.)

Crow, J. F., and M. Kimura. 1970. *An Introduction to Population Genetics Theory*. Harper and Row. (More advanced.)

Kimura, M., and T. Ohta. 1971. *Theoretical Aspects of Population Genetics.* Princeton University Press.

Li, C. C. 1955. *Population Genetics*. University of Chicago Press. (Unusually clear and easy to understand.)

CHAPTER 23, 24

Falconer, D. 1960. *Quantitative Genetics.* Ronald Press. (I think this is the best all-round book for quantitative genetics. Emphasis on animal breeding.)

CHAPTER 25

Dobzhansky, T. 1970. *Genetics of the Evolutionary Process*. Columbia University Press. (Beautifully written.)

Fisher, R. A. 1958. *The Genetical Theory of Natural Selection.* Dover. (Hard to read, but this is the Bible of selection theory.)

Haldane, J. B. S. 1932. *The Causes of Evolution.* Dover. Reprinted as a paperback. (Still worth reading after all these years.)

Simpson, G. G. 1953. *The Major Features of Evolution.* Columbia University Press.

CHAPTER 26

Horowitz, N. H., F. D. Drake, S. L. Miller, L. E. Orgel, and C. Sagan. 1970. In *Biology and the Future of Man*, P. Handler (ed.), Oxford University Press.

Orgel, L. E. 1973. *The Origins of Life.* John Wiley.

STATISTICAL APPENDIX

Fisher, R. A. *Statistical Methods for Research Workers.* Oliver and Boyd. (Many editions. The standard book for many years.)

Mather, K. 1965. *Statistical Analysis in Biology.* University Paperbacks, London.

Snedecor, G. W., and W. G. Cochran. 1967. *Statistical Methods.* 6th ed. Iowa State College Press. (practical, how-to-do-it statistical methods.)

INDEX

Aberrations, chromosomal, 121ff
ABO blood groups, 151, 163
Acentric chromosome, 126
Acquired character, 213
Adaptive enzyme, 146
Addition rule, 40
Additive genes, 193ff
Adenine, 12, 96
Agglutination, 151
Albinism, 83, 91, 141
Alkaptonuria, 91
Alleles, 4, 111
Allelism, test for, 114
Allopolyploidy, 122
Altruism, selection of, 122
Amino acids, 15, 90, 92, 95ff, 105ff, 218ff, 226
Amniocentesis, 48
Amorph, 113
Amphidiploid (= allotetraploid), 122
Analysis of variance, 204, 239ff
Anaphase, 20
Aneuploidy, 23ff, 123ff
Antibodies, 151, 157
  variability of, 157
Antigens, 151
Antimorph, 113
Antimutagen, 136
Antiserum, 151
Arginine, 90
Ascospores, 53, 89
Ascus, 53, 89
Assortative mating, 161
Attached-X chromosomes, 33
Auerbach, C., 135
Autogamy, 54
Autonomous development, 32ff
Autopolyploidy, 122
Autosome, 27
Autotroph, 54, 225

Backcross, 3
Bacteria
  genetics of, 69ff
  life cycle of, 54
  size, 55
Bacteriophage, 55
  life cycle of, 55, 69ff
  mapping, 115
Baltimore, D., 95
Bar eye in Drosophila, 125
Barr body, 34
Bayes' formula, 49
Beadle, G. W., 89
Bees, 27
Benzer, S., 115
Binomial distribution, 41
Biochemical genetics, 89ff
Blending inheritance, 1, 214
Blood group mosaics, 156
Blood groups, 151ff, 163, 191
Blood transfusion, 152
Bridges, C. B., 30

Cairns, J., 16
Cell, 19ff
Centromere, 20, 126
  mapping, 62
Centrosome, 20, 64
Chemostat, 135
Chiasma, 64ff
Chi-square test, 229ff
  chart of, 241
Chromosome, 19
  aberrations, 121ff
  mapping, 57ff
  number, 19
  polytene, 121
  sex, 27ff
  theory of heredity, 23
Cistron, 117
Citrulline, 90
CO_2 susceptibility in Drosophila, 142
Code, genetic, 105ff
Codon, 100, 105ff
Coefficient of inbreeding, 170ff
Coefficient of kinship (consanguinity), 172
Coiling in snails, 79
Coincidence, 60ff
Colchicine, 122
Color blindness, 29, 165
Comb in fowls, 6
Complementation, 114, 117, 131
Concordance, 81
Conditional probability, 40
Confidence limits, 235, 240
Conjugation, 54
Consanguineous mating, 169ff
Constitutive enzyme, 146
Continuous variation, 193ff, 203ff
Corn, hybrid, 175
  life cycle of, 52
Correlation, 205
Correns, C., 1
Counseling, 47
Cousin marriage, 174
Covariance, 205
Creeper fowl, 57
Crick, F. A. C., 12, 105, 108
Crossing-over, 57ff, 126
Crossover suppressor, 126
Cytochrome, evolution rate, 219
Cytoplasm, 19, 141
Cytoplasmic inheritance, 141ff
Cytosine, 12, 96

Darwin, C., 213, 225
DDT, resistance to, 196, 221
Deficiency, 120, 137
Degrees of freedom, 204, 233ff
Delayed gene expression, 79
Deletion, 120, 137
Deoxyribonucleic acid, 11, 95ff
De Vries, H., 1, 214
Diad, 22, 64ff
Dicentric chromosome, 126
Dihybrid cross, 5ff
Diploid, 21
Diplotene, 64
Dizygotic twins, 81

DNA, 11, 95ff
  length of, 26
Dominance, 2, 82, 112
Down's syndrome, 24
Drift, random, 176
Drosophila:
  absence of crossing-over in male, 61
  attached chromosomes in, 33
  gene number in, 131
  sex determination in, 28ff, 130
Duplication, 120, 216

Endosperm, 55
Environment, 77ff
  variance due to, 204ff
Enzymes, 89ff, 146ff
  regulation of, 146
Episome, 70ff, 147
Epistasis, 82, 205ff
Equilibrium, gene frequency, 185ff
Erythroblastosis foetalis, 154
Escherichia coli, 54, 69ff, 137ff
Euchromatin, 129
Eugenics, 200
Eukaryote, 25
Euploidy, 122
Euthenics, 200
Evolution, 213ff
  molecular, 216ff
Exobiology, 228
Expressivity, 79
Extranuclear inheritance, 141ff

F (inbreeding coefficient), 170ff
  episome in *E. coli,* 72ff
F_1, F_2, 2
F′ episome, 93, 147
Factor (gene), 2
Factorial, n!, 41
Feedback inhibition, 145
Fibrinopeptide, evolution rate of, 219
First division segregation, 62
Fisher, R. A., 11, 153, 154, 203, 235
Fluorescent chromosome staining, 31
Founder effect, 177
Frame-shift mutation, 105, 137
Fraternal twins, 81
Freemartin, 156

Galactosemia, 92
Gamete, 2ff
Gametophyte, 21
Garrod, A., 89
Gene, 2, 117
  number, 131
Genetic code, 105ff
Genetic counseling, 47
Genetic prognosis, 67
Genic balance, 123, 130
Genotype, 4
Grandchildless, in Drosophila, 80
Guanine, 12, 96
Gynandromorph, 31

Haldane, J. B. S., 225
Haploid, 21, 122
Haptoglobin, 217
Hardy-Weinberg rule, 162ff
Hemizygous, 29
Hemoglobin, 46, 102, 188, 217
Heritability, 198, 207ff
Heterochromatin, 129
Heterosis, 174ff, 188
Heterotroph, 54, 225
  theory of origin of life, 225
Heterozygote advantage, 188
Heterozygous, 4
HFR, 72
Histocompatibility, 156
Histone, 25, 121, 219
  evolution rate, 219
Homozygosity, effect of inbreeding on, 169
Homozygote, 4
Hormone, 33
Horowitz, N., 228
Hybrid vigor, 174ff
Hymenoptera, sex determination of, 28
Hypermorph, 113
Hyphae, 53
Hypomorph, 113
Hypostasis, 82

Identical twins, 81
Immunogenetics, 141ff
Immunoglobulin, 157
Inborn errors of metabolism, 89ff
Inbreeding, 169ff
Inbreeding coefficient, 170ff
Independent assortment, 5
Inducer, 146, 148
Industrial melanism, 221
Ingram, V., 102
Initiation, transcription, 98, 146ff
  translation, 107
Interaction, gene, 82, 206
Interference, 60
Interphase, 20
Intersex, 34, 130, 156
Inversion, 125ff
  effect on crossing-over of, 126
Isozyme, 92, 220

Jacob, F., 72, 146
Jukes, T., 221

Kallmann, F., 81
Kalmus, H., 81
Kappa, in Paramecium, 142
Karpechenko, G., 122
Khorana, H. G., 15, 106
Kimura, M., 189, 221
King, J., 221
Kinship, coefficient of, 172
Klinefelter's syndrome, 34
Kosambi, D. C., 61

Lamarck, J., 213
Lambda phage, 69
Landsteiner, K., 151ff
Lederberg, J., 54, 70
Leptotene, 64
Lethal equivalents, 174
Lethal genes, 77, 131, 134
Levine, P., 153, 155
Life cycles, 51
  bacteriophage, 55, 69
  corn, 52
  Escherichia coli, 54, 69
  Neurospora, 53
  Paramecium, 54
  Sphaerocarpos, a liverwort, 52
Likelihood ratio, 48
Linkage, 57ff
  sex-, 28
Liverwort, life cycle of, 52
Lyon, M., 34
Lysogeny, 69

Maize, life cycle of, 52
Malaria, 189
Male sterility in maize, 141
Mapping function, 61
Mapping of chromosomes, 58ff, 70ff, 115, 121
  centromere, 62
Maternal influence, 79
McKusick, V., 47
Mean, 203, 239
Meiosis, 21, 61ff
Melandrium, sex determination in, 130
Melanism, industrial, 221
Mendel, G., 1ff
Mendelian inheritance, 1ff
  evolutionary advantage of, 222
Meselson, M., 15

Messenger RNA, 97ff, 105ff, 146ff
Metaphase, 20, 64
Micronucleus, Paramecium, 54
Miller, S., 226
Mitochondria, 142
Mitosis, 20ff
M-N blood groups, 153, 164
Modifying genes, 83
Mongolism, 24, 123, 128
Monod, J., 146
Monohybrid cross, 1
Monosomic, 23, 34, 123
Monozygotic twins, 81
Morgan, T. H., 28ff, 58ff
Muller-5 method, 134
Muller, H. J., 74, 113, 134ff, 201, 223
Multinomial distribution, 42
Multiple alleles, 111ff
Multiple factors, 193ff
Multiplication rule, 40
Mutation, 7, 14, 133ff
  chemical mutagens and, 135
  evolution and, 214ff
  molecular basis of, 136
  rate, 8, 133ff
  repair, 138
  temperature effects on, 135
  Watson-Crick model and, 14
  X-ray induction of, 135
Muton, 117
Myoglobin, 217

Natural selection, 181ff, 213ff, 225ff
Neomorph, 113
Neurospora, 53, 89
Neutral evolution theory, 189, 221
Newman, H. H., 82
Nirenberg, M., 105
Nondisjunction, 23, 30, 123
Novick, A., 135
Nuclease, 138
Nucleic acid, 11ff, 95ff, 105ff, 225ff
Nucleoprotein, 25, 121
Nucleoside, 96
Nucleotide, 96
Nucleus, 19

Odds, 39, 49
  prior and posterior, 49
Oocyte, 21, 65
Operator, 147
Operon, 146ff
Origin of life, 225
Origin of species, 223
Ornithine, 90
Overdominance, 175, 188
Owen, R., 156

Paramecium, 54, 142
Partial transfer of genes in *E. coli*, 72ff
Particulate inheritance, 1, 214
Peas, 1ff
Pedigrees, human, 45
Penetrance, 79
Phenotype, 4
Phenylalanine, 90
Phenylketonuria, 91
Phenylpyruvic acid, 91
Plastid inheritance, 141
Pleiotropy, 78
Pneumococcus, transformation with, 73
Polar body, 21
Polygenic inheritance, 193ff, 203ff
Polymerase, 95ff, 138
Polymorphism, 78, 188
Polyploidy, 122
Polysomy, 123
Polytene chromosomes, 121
Population genetics, 161ff
Position effect, 115, 129
Probability, 39ff, 229ff
  conditional, 40
  posterior, 48
Progeny test, 199
Prokaryote, 25
Promoter, 147
Prophage, 69ff
Prophase, 20
Propositus, 45
Protein, 15, 95ff, 216ff
Punnett square, 6
Purine, 12, 96
Pyrimidine, 12, 96

Q banding, 31
Quantitative traits, 193ff

Random gene frequency drift, 175ff, 189, 221
Random mating, 161ff
Raphanobrassica, 122
Recessive, 2
Recombination, 5, 57ff, 69ff,
  intragenic, 63, 115ff
Recon, 117
Regression, 117, 198, 203ff
Regulator genes, 145
Replication, 14, 153
Repressor, 142, 146ff
Restriction enzyme, 74
Reverse transcriptase, 95
R factors, 71, 196
Rh blood groups, 153, 191
Ribonucleic acid, 95
  messenger RNA, 95ff, 105ff
  ribosomal RNA, 100
  transfer RNA, 99, 105ff
Ribose, 95
Ribosome, 100
RNA, 95, 100, 105ff
Roan cattle, 5
Roentgen, or R, 135

Salivary gland chromosomes, 121
Second division segregation, 62
Segregation, 2ff
Selection, 181ff
Selection index, 200
Self-fertilization, 2, 111, 169
Self-sterility genes, 111
Semi-sterility, 127ff
Sex chromosomes, 27ff
Sex determination, 27, 130
Sex differentiation, 33
Sex-linkage, 28ff
Sex reversal, 33
Sexual reproduction, evolutionary advantage of, 51, 222
Sib mating, 170
Sickle cell hemoglobin, 46, 102, 188
Significance test, 229ff
Silk moth, 34
Single active X principle, 34
Snails, direction of coiling in, 79
Soluble RNA (see Transfer RNA)
Sonneborn, T., 142
Spermatid, 21ff
Spermatocyte, 21ff
Spermatozoa, 21ff
Sphaerocarpos, life cycle of, 52
Spirochaete in Drosophila, 143
Sporophyte, 21
Stahl, F., 15
Standard deviation, 204, 236, 240
Statistical methods, 203ff, 229ff
Streptomycin, 100
Structural gene, 146ff
Sturtevant, A. H., 58, 127
Super female, 30
Suppressor genes, 83, 101
Sutton, W., 23
Synonymy, 107

Tatum, E. L., 54, 88
Tay-Sachs disease, 48, 77, 92
Telophase, 20
Temin, H., 95
Temperate phage, 69
Testcross, 3
Tetrad, 64ff

Tetraploid, 122
Thymine, 12, 96
Time scale, evolutionary, 215
Tissue transplants, 156
Tolerance, immune, 156
Transcription, 15, 95, 97
Transduction, 70ff
Transfer RNA, 99, 105ff
Transformation, 73
Transition mutations, 137
Translation, 15, 95, 99
Translocation, 127
Transversion mutations, 137
Triploid, 87
Trisomic, 24, 123
Tritium, 16
Turner's syndrome, 34
Twins, 81
Tyrosine, 90
Ultraviolet radiation, 117, 137, 138
Uracil, 96
Urea cycle, 90

Variance, 203ff, 239ff
  analysis of, 204, 239ff
  environmental, 205
  genetic, 205
Variegation, 129
Vinograd, J., 15
Viruses, 55, 69
Von Tschermak, E., 1

Watson-Crick model of DNA, 12, 96
Watson, J., 12
Weismann, A., 214
Wiener, A., 153
Wild-type, 78
Wobble hypothesis, 108
Wollman, E., 72
Wright, S., 170, 203

X chromosome, 27ff
Xeraderma pigmentosum, 117, 138
X-linkage, 28ff
X-radiation, 135

Y chromosome, 27
Y-linkage, 29

Zinder, N., 70
Zygote, 4, 21
Zygotene, 64

(I am indebted to Dr. P. S. Dawson, who prepared an index to an earlier edition from which this is largely taken. J. F. C.)

Burgess Publishing Company

Burgess Publishing Company